Year : 2012

The Normative Dimension of Social Life from a Second-Person Perspective

Del Mar Maksymilian

UNIL | Université de Lausanne

FACULTÉ DES SCIENCES SOCIALES ET POLITIQUES

The Normative Dimension of Social Life from a Second-Person Perspective

THÈSE DE DOCTORAT

présentée à la

Faculté des Sciences Sociales et Politiques
de l'Université de Lausanne

pour l'obtention du grade de

Docteur ès sciences sociales

par

Maksymilian Del Mar

Directrice de thèse

Mme Laurence Kaufmann, Professeure, Université de
Lausanne

Membres du jury

Mme Michèle Grossen, Professeure, Université de Lausanne
M. Julien Deonna, Professeur, Université de Genève
M. Paul A. Roth, Professeur, Université de California-Santa-
Cruz

LAUSANNE

2012

UNIL | Université de Lausanne

Faculté des sciences
sociales et politiques

IMPRIMATUR

Le Conseil de la Faculté des sciences sociales et politiques de l'Université de Lausanne, sur proposition d'un jury formé des professeurs

- Laurence KAUFMANN, directrice de thèse, Professeure à l'Université de Lausanne

- Michèle GROSSEN, Professeure à l'Université de Lausanne

- Julien DEONNA, Professeur à l'Université de Genève

- Paul A. ROTH, Professeur à l'Université de California-Santa Cruz

autorise, sans se prononcer sur les opinions du candidat, l'impression de la thèse de Monsieur Maksymilian DEL MAR, intitulée :

« *The Normative Dimension of Social Life from a Second-Person Perspective* »

Lausanne, le 12 mars 2012

Le Doyen de la Faculté

Professeur
René Knüsel

2

Abstract

This thesis argues that insofar as we want to account for the normative dimension of social life, we must be careful to avoid construing that normative dimension in such a way as to exclude that which the second-person perspective reveals is important to social life and our ability to participate in it.

The second-person perspective reveals that social life ought to be understood as a mix or balance of the regular and the irregular, where, in addition, those one interacts with are always to some extent experienced as other in a way that is neither immediately, nor perhaps ultimately, understandable. For persons to be able to participate in social life, conceived of in this way, they must have abilities that allow them to be, to some extent, hesitant and tentative in their relations with others, and thus tolerant of ambiguity, uncertainty and unpredictability, and responsive to and capable of learning from the otherness of others in the course of interacting with them.

Incorporating the second-person perspective means we have to make some changes to the way we think about the normative in general, and the normative dimension of social life in particular. It does not mean giving up on the distinction between the normative and the regular – that continues to be fundamentally important – but it does mean not excluding, as part of social life and as worthy of explanation, all that which is irregular. A radical way of putting it would be to say that there must be a sense in which the irregular is part of the normative. A less radical way, and the way adopted by this thesis, is to say that any account of the normative dimension of social life must not be such as to exclude the importance of irregularity from social life. This will mean 1) not characterising conventions, norms and rules as determinants of appropriateness and inappropriateness; 2) not thinking of them as necessary; 3) not thinking of them as necessarily governing minds; and 4) not thinking of them as necessarily shared.

Résumé

L'argument principal de la thèse est que, pour rendre compte de la dimension normative de la vie sociale, il faut veiller à ne pas exclure la perspective de la deuxième personne – une perspective importante pour comprendre la vie sociale et la capacité requise pour y participer.

Cette perspective nous permet d'imaginer la vie sociale comme un mélange ou un équilibre entre le régulier et l'irrégulier, l'interaction entre des individus pouvant être appréhendée comme l'expérience de chaque personne avec «l'autre» d'une manière qui n'est pas immédiatement compréhensible, et qui ne peut pas, peut-être, être ultimement comprise. Pour participer à la vie sociale, l'on doit avoir la capacité de rester hésitant et «réactif» dans ses relations avec les autres, de rester ouvert à leur altérité et de tolérer l'ambiguïté, l'incertitude et l'imprévisibilité des interactions sociales.

Adopter une perspective «à la deuxième personne» conduit à une autre manière de penser la normativité en général, et la dimension normative de la vie sociale en particulier. Cela ne veut pas dire qu'il faut abandonner la distinction entre le normatif et le régulier – une distinction qui garde une importance fondamentale – mais qu'il faut reconnaître l'irrégulier comme faisant partie de la vie sociale et comme étant digne, en tant que tel, d'être expliqué. Une conception radicale pourrait même concevoir l'irrégulier comme faisant partie intégrante de la normativité. Une approche moins radicale, qui est celle adoptée dans cette thèse, est de dire que tout compte-rendu de la dimension normative de la vie sociale doit prendre en considération l'importance de l'irrégularité dans la vie sociale. Une telle approche implique que les conventions, normes et règles (1) ne déterminent pas ce qui est approprié ou inapproprié; (2) ne sont pas toujours nécessaires ; (3) ne gouvernent pas le fonctionnement de l'esprit ; et (4) ne sont pas nécessairement partagées.

Acknowledgements

Contents

Epigraph

Good Manners and Foolish Khushim

Until he was sixteen, Khushim the Fool lolled about on top of the oven and never went anywhere. His mother said, 'Khushim, why don't you get out of the house? It'll do you good to rub shoulders with other people.'

'All right, mother,' he said. So he left the house and went out in the street, and there he saw a crowd. He pressed his way into the middle of it and started to rub up against the people, for which they beat him soundly. When he came home weeping, his mother said, 'My son, when you mingle with people you have to greet them. If it's morning, you say, "Good morning." If it's afternoon, you say, "God be with you." If it's evening, you say, "Good evening." When you take your leave in the afternoon, you say "Good day," and if it's night you say "Good night".'

'Good, Mother,' said Khushim, 'I'll do as you say.'

He went out into the street, and there he saw a funeral procession. He ran up to the mourners and cried, 'Good morning. God be with you. Good evening. Good day. Good night,' for which they beat him soundly.

When he came home weeping, his mother said, 'My son, if you come upon a funeral procession, you're supposed to weep. And if the deceased is a young man, you say, "Woe, woe. Such a young tree to be felled in its prime." And if he's old, you say, "May he rest in paradise".'

'Good, Mother,' said Khushim, 'I'll do as you say.'

He went out into the street, where he saw a wedding procession going by. The musicians were playing, the relatives of the bride and groom were dancing. Khushim, seeing them, cried out, 'Woe, woe. Such a young tree to be felled in its prime. May he rest in paradise,' for which they beat him soundly.

When he came home weeping, his mother said, 'My son, if you see a wedding process, you cry, "Congratulations!" and you sing and dance.

Khushim said, 'Good, Mother. I'll do as you say.'

He went out into the street and came to a house on fire. People were standing around it weeping and wailing. Khushim called out 'Congratulations!' and began to sing and dance, for which they beat him soundly.

When he came home weeping, his mother said. 'My son, if you see a house on fire you're supposed to grab a bucket of water and put out the flames.'

Khushim said, 'Good, Mother. I'll do as you say.'

He went out into the street, and there he saw a house with smoke coming out of the chimney. So he grabbed a bucket of water, climbed up on the roof, and poured water down the chimney, for which they beat him soundly.

When he came home weeping, his mother said, 'My son, you may as well loll about on top of the oven. There's no point in your rubbing shoulders with the rest of the world.'

(Folktale #76, Weinrich 1988, 216-7)

Introduction

A. Preface and Overview

There are two ways of reading why Khushim – the hero or anti-hero of the epigraph to this thesis – does not get along with others. The first of these ways is to argue that Khushim does not have enough knowledge of the conventions, norms and / or rules that govern the community he wishes to participate in. This is a problem for Khushim, for those conventions, norms and / or rules determine what it is appropriate or inappropriate to do that in that community. It follows that because Khushim has inadequate knowledge of those 'determinants of appropriateness and inappropriateness', which everyone else in the community shares, so he is bound to fail to get along with others. Perhaps, then, he is indeed better off, as his mother painfully concedes, lolling about on top of the oven. Khushim's knowledge of those determinants is inadequate in numerous respects: he does not, for instance, know enough of the determinants – he just knows the one that his mother tells him immediately before venturing out; further, he has virtually no knowledge of the conditions of applicability of those determinants, i.e. when they apply and when they do not apply. The claim, based on the first reading of why Khushim fails, is that if he had this knowledge – and, better still, if he internalised it such that it became second nature to him, functioning through him without any deliberative effort on his part, for instance in the form of affective-corporal dispositions – then Khushim would have been fine, and his mother very proud of him.

That is the first reading of the tale, but there is a second possibility. This second reading suggests that what Khushim is missing is not knowledge concerning the conventions, norms and / or rules that – from the perspective of this reading – are better seen as contingently useful resources (rather than determinants of appropriateness and inappropriateness). Instead, what Khushim is missing are certain basic abilities to get along with others. Notice, for instance, Khushim's lack of patience: he rushes in, barely considering the relevance of the normative resource he has been given by his mother, following it blindly. Khushim does not hesitate; he is not, at any point, tentative in his relation with others. Indeed, what is most striking about his adventure is that he hardly relates to others at all – it would be more accurate to say that he treats them as if they were characters in a script governed by the one normative resource that he has at the forefront of his mind. Khushim has, it

appears, no toleration for ambiguity in his relation with others; no sense that they could be different in ways that he might not, immediately or even ultimately, understand; no willingness to be surprised by others and respond to them in a way that is novel for him; indeed, no ability to learn from others. He is, instead, very assertive and strongly committed to a particular normative resource that he does not consider might be only contingently useful. His world is a world of perfect order, in which there is no scope for irregularity, unpredictability and uncertainty.

This thesis is more sympathetic to the second reading. Put differently, it is sympathetic to the idea that there are certain basic abilities that persons need to get along with each other, and that theorists who wish to try to account for how persons manage to get along need to make sure they make room for, and carefully describe, these abilities. As the above reading suggests, the thesis argues that those basic abilities have something to do with being tolerant of ambiguity, with being able and willing to experience irregularity, with being hesitant and tentative in one's relation with others.

The second reading is not – it has to be emphasised – some naïve, romantic idealisation. It is not about persons being nice or kind to each other. But, of course, the second reading is making a positive claim: it is saying that these abilities are needed for persons to be able to get along with each other, at least a lot of the time. To get along with each other – and, more broadly, to participate in social life – persons need to be responsive to each other. Of course, they can be responsive in ways that others will find obnoxious, rude and offensive. Their responsiveness might land them in deep trouble with others. However, to be able to relate to others at all, and, arguably, to get along with them most of the time, persons do need a set of abilities, especially those that, as noted above, enable a person to change and adapt in the course of interacting with others, and thus learn from others who are, and must always remain to some degree, different to them.

It is also important to emphasise, at the outset of this thesis, that the second reading does not dismiss the value of normative resources. On the contrary, it builds them in as potentially useful guides for what one ought to do and standards for how to legitimately evaluate others. It does not, however, treat those resources as determinants of appropriateness and inappropriateness that are necessary for that community; nor does it say that the minds of those in the community are necessarily governed by them; and nor does it say that those determinants are necessarily shared.

Nevertheless, it allows for persons to experience, in certain contexts and for certain purposes, what normative resources recommend, require or demand (depending on their degree of prescriptiveness) as unchangeable; in that sense, it allows persons to experience normative resources as necessary. Apart from building in this possibility of experiencing normative resources as necessary, it replaces necessity as it appears in the first reading with contingency. Conventions, norms and rules are not, it suggests, necessary – they are only ever contingently useful. Further, they do not necessarily govern the minds of any one community, such that the cognition of those members of the community can be described as either necessarily or by default following those normative resources. And, finally, the second reading does not agree that it is necessary for members of the community to share knowledge of (including commitment to) those same normative resource – it acknowledges that, on occasion, they may share them, but it insists that it is not necessary that they do so.

What has been referred to above as the 'first reading' is, roughly, what is called in this thesis 'the governance view of the normative dimension of social life'. The governance view has four features: first, it characterises conventions, norms and / or rules as determinants of appropriateness and inappropriateness; second, it argues that those determinants of appropriateness and inappropriateness are necessary – either metaphysically, as is claimed sometimes for rules of logic, or practically, as is claimed sometimes for the rules of games; third, it argues that those determinants necessarily govern the minds of persons who are members of the relevant community or domain (where the relevant rules are necessary); and fourth, it argues that the members of the relevant community or domain necessarily share those determinants. Strictly speaking, only the latter three are arguments, and in this thesis they are treated as the 'three elements' of the governance view. The first is, as it were, a presupposition – for the three elements to hang together, conventions, norms and / or rules must be characterised as determinants of appropriateness and inappropriateness.

The governance view has important implications for 1) our understanding of the social; and 2) following on from that, our understanding of the abilities persons need to participate in the social. Under the guise of capturing the normative dimension of social life, then, the governance view greatly narrows down our conception of the social in general. Put differently, it sets up a particular target or object of explanation: social order. The setting up of this target or object of explanation can occur in two different ways: first, either a set of regularities is said to

have been observed, and then, in order to build in a normative dimension, a set of normative attitudes are built on top of those regularly performed behaviours – this is the move that characterises, on the way that these terms are rationally reconstructed in this thesis, an order composed of conventions and norms; or, second, a set of rules (conceived as linguistically-expressed normative propositions issued by authorities) is thought to be necessary for some community or domain, and thereafter, regularised, i.e. persons are imagined to either consciously or unconsciously acquire and internalise those rules, and develop abilities, skills or dispositions to follow them. In effect, both ways end up with the same coincidence of regularity and normativity; both, in that sense, end up with a target or object of explanation that captures the normative dimension of social life as an order or domain of patterns that are at once regular and normative.

The effect of this conception of the social on the description of the abilities persons need to participate in social life is profound. As already indicated above, on the governance view, social minds are described as if they were in essence convention-, norm- or rule-following things. The key problem that occurs, again and again, is how to conceive of social minds as such convention-, norm-, or rule-following things without requiring those minds, every time they act, to compute – or consciously deliberate about the applicability of – those conventions, norms or rules. A related problem is explaining how social minds can come to be committed to those conventions, norms or rules, especially in the context of the norms and rules, which typically – at least on the rational reconstruction offered here – arise in order to limit or control non-social (selfish) interests. The solution is often claimed to exist in such versions as 'normative dispositions' – dispositions, affective and corporeal in character (embodied emotions, one might say), to follow, and at once be committed to, those conventions, norms or rules. Following these dispositions is often said to have two prongs: first, a disposition to use them as guides for one's own behaviour; and second, a disposition to use them as standards to evaluate (and thus discipline) both one's own behaviour and the behaviour of others.

This thesis argues against the governance view of the normative dimension of social life precisely because of its narrowing of our understanding of the social and because of the related effects on how the abilities of persons to participate in the social are described. At its most general, then, what is philosophically at stake in the thesis is precisely our understanding of the social and our abilities to participate in it.

In order not to narrow down, from the beginning, what we are setting out to understand, we must avoid adopting the governance view – we must dismantle it, and show that we can do without it. This does not mean, however, that we need to throw the baby out with the bathwater, i.e. there are aspects of the governance view that are extremely important for our understanding of the social and participation in it. No understanding of the social, including participation in it, can proceed without reference to normative resources, especially to normative attitudes, such as those exemplified by talk of conventions and norms, and those linguistically-expressed normative propositions issued by authorities that, in this thesis, we refer to as rules. The point, however, is not to begin with those normative resources and thereby allow them to set up the target or object of explanation, with all the consequences this has – as noted above – on describing our abilities to participate in social life, including having normative experiences in the course of interacting with others.

Put this way, it might seem as if the thesis is wholly negative. This is not the case. Certainly, a large part of this thesis is devoted to examining – including faithfully articulating – and ultimately critiquing the governance view. Thus, and as will be elaborated on a little more in the next Section of this Introduction – each of the first four Parts of the thesis is devoted to an aspect of the governance view. However, in the course of examining and critiquing the governance view, a positive picture is developed alongside it. That positive picture is then systematised and articulated in the fifth and sixth Parts of this thesis.

The key to the positive picture this thesis develops is the second-person perspective. More will be said below – in the third Section of this Introduction – as to how that term is used here in ways that partly overlaps, but is also distinguishable, from previous uses of that term in a variety of literatures (including philosophical, sociological and psychological). For the moment, we can say this: the second-person perspective is a perspective on social life in general. It is a way of seeing what ought to be part of any target or object of an explanation of the social, and thus also what needs to be described if we are setting out to understand the social. The theorist imagines what a social agent, positioned vis-à-vis another agent – typically in a face-to-face relation – needs to be able to do in order to get along with that person. The theorist imagines what experiences that agent needs to have. The theorist also imagines what the interaction needs to be like in order to be going well and not breaking down. The second-person perspective is, then, a looking-glass onto social

life: it allows us, as theorists, to pick out what is important about social life that we might otherwise miss. It thereby also indicates what needs to be described, i.e. in this case, what kinds of abilities persons need in order to be able to participate in interaction understood from the second-person perspective. In that sense, the second-second perspective might be thought to represent a call for a *gestalt-shift* – a change in the way we imagine what is part of social life, and, consequently, in the way we describe our abilities to participate in it.[1]

What, then, does the second-person perspective on social life pick out as worthy of explanation and as requiring description? The short answer is that it picks out all that which is irregular, uncertain and unpredictable. Its point, however, is not to replace regularity with irregularity, or certainty with ambiguity, or predictability with unpredictability, but to make a plea for getting over that divide or dichotomy between meaningful order on the one hand and meaningless chaos on the other, and to recognise, instead, a mix or balance of the regular and the irregular. The point, then, is to broaden the scope of our understanding of the social as always including some regularity and some irregularity; some reoccurring patterns and some contingent emergence of novel relevancies; some predictability and some unpredictability. The second-person perspective recommends, if you like, a change in the object or target of explanation. What we are seeking to explain, as theorists of the social, is not order, or a realm of patterns that are at once regular and normative. Rather, we are seeking to explain the mix or balance of the regular and the irregular, the patterned and un-patterned, the ordered and dis-ordered – but with the proviso that we understand that the plea of the second-person perspective is not to get too overwhelmed by those distinctions. For the sake of making its positive claim, and its argument more striking, the second-person perspective emphasises – it seeks out – that which is irregular in social life. But, once again, it should not be thought of as limited to that irregularity; again, it instead emphasis a mix or balance between the regular and the irregular.

It might be thought that the second-person perspective is plausible as a conception of the social, but not of its normative dimension. There is some truth in that scepticism. It is not easy to imagine how else we could understand the normative except as a domain of conventions, norms and rules that determine what is

---

[1] Note that sometimes in the sources discussed throughout this thesis reference to perspectives – whether first-personal, second-personal or third-personal – is meant in another sense, i.e. the sense of an agent adopting a certain perspective in relating with someone.

appropriate and inappropriate to do. The normative has to be understood as somehow distinct from the regular, and it has often been thought that that job can only be done by conventions, norms and / or rules – some phenomena that prescribe what ought to be done, and determine what is appropriate / correct or inappropriate / incorrect. Strictly speaking, this thesis does not argue against this approach to the normative, although it does counsel taking care as to what role one ascribes to conventions, norms or rules. The argument is that if we take the second-person perspective on social life seriously then we need be very careful about what we say about the normative in general, but especially about the normative dimension of the social. The second-person perspective on social life tells us that insofar as we want to account for the normative dimension of social life, we ought not to do so in such a way that will exclude by definition, or squeeze out the very possibility of, the irregularity that is there in interaction, including all the abilities we have as participants in social life that enable us to be responsive to, and learn from, that irregularity. Another way of putting it is that from the second-person perspective what becomes necessary is that we qualify or soften the governance view; that we amend it so that it does not make it impossible to see what the second-person perspective allows us to see about social life and our ability to participate in it. This includes recognising that persons have, and need to have, normative experiences that are hesitant and tentative, e.g. appreciating that this or that might perhaps be appropriate or inappropriate.

The second-person's perspective of the social has, or so it is argued in this thesis, important implications for how we describe the abilities that persons need to participate in the social. We can no longer describe – as the governance view would have us describe – social minds as convention-, norm- or rule-following things. If social minds were just those kinds of things then they would be unable to experience, let alone be responsive to, irregularity. They would thereby be bereft of one of the key – indeed basic – abilities they need to participate in social life. Again, social life, or any one interaction, always has a mix of the regular and the irregular. As a result of the presence of the irregular, something novel – unpredictable, ambiguous, risky, and uncertain – can emerge as relevant for that interaction, including relevant for experiences about what it might be appropriate or inappropriate to do in the course of that interaction. The regular and the irregular can co-exist side by side: persons can, to some extent, take certain normative resources as contingently useful and use them to

navigate the regular, but without abilities that allow them to be responsive to irregularity, they will – like Khushim did – struggle to get along with others.

It is also important to note that the second-person perspective on social life picks out as vital to that life the recognition of the otherness of others. This insistence on the otherness of others, as part of the second-person perspective, renders it dangerously close to some form of idealisation or even moralisation. This thesis does what it can to distance itself from that interpretation of this feature of the second-person perspective, but it is difficult to divorce the idea completely from it. Perhaps the best way of articulating this feature of social life from the second-person perspective is to say that it wants to emphasise that other persons in social life cannot be experienced – if one is to relate with them, including get along with them – as ever being wholly, and certainly never immediately, understandable. If one relates with another person – if one interacts with them – then one will always the experience the other as unfinished and incomplete; in short, as different in a way that one cannot quite grasp. In that sense, an interaction with another person is always marked by some degree of surprise; a sense in which the other, as a source of her own autonomy, can change the direction and flavour of the interaction in a way we cannot immediately, or perhaps ever ultimately, understand, but which we – if we are to interact with them – must respond to. The other is always a source of invitations and opportunities to interact that we could not have been prepared for. If there is not this uncertainty in one's interaction with another person, then, arguably, one is not interacting or relating with them; instead, one is treating them as if they were a puppet in one's control, or a character in a script that one is the sole author of. In that sense, too, interactions or relations with others can never be completely fluent and easy – there is always some element of difficulty. Of course, that is not a down-side to interaction – it is one of its most positive features. As difficult as it is to articulate, it is, together with the mix or balance of regularity and irregularity, a fundamental feature of the second-person perspective on social life – indeed, one can also understand it as further helping account for the presence of irregularity in interactions with others.

In this thesis, the abilities that persons need to participate in social life, when that social life is understood from the second-person perspective, are described as a partnership between 'appreciation' and 'attentive silence'. These are technical terms in this thesis, and they are explained in some detail in the sixth and final Part. They

have some similarities: both are tentative and hesitant forms of relating with others (or, more generally, perhaps also with the environment). They enable, precisely, tolerance of ambiguity; they are responsive to and allow for learning from the unpredictable. They are both key ways in which persons are able to get along with others that are different to them; to respond to, but also to take pleasure in, the otherness of others.

But there are, also, some differences. Appreciation, in brief, refers to the ability to appreciate that someone, including something someone does, might perhaps be such-and-such, e.g. elegant or sad, or appropriate or inappropriate. It is best distinguished from the activity of justification or systematisation: in other words, whereas justification or systematisation depend upon the identification of properties that then constitute the grounds for the making of certain (justifiable or systematisable) claims, appreciation operates without the prior identification of such properties and is, in that sense, groundless. It being groundless, however, does not render it solipsistic; on the contrary, it is inherently more flexible, more revisable, and thus more open to correction than justification and systematisation. It is, in short, a very hesitant and tentative way of giving form to one's experience of the external world, including others. Attentive silence, on the other hand, does not involve giving form to one's experience. Where appreciation can be understood to be a way of being passively active (active because one is giving form to one's experience), attentive silence could be seen, instead, to be a very active way of being passive. When we attend silently to something, or someone, we allow ourselves to be infused with the experience of it or him / her. This cognitive ability can best be illustrated with an example: when running one's hand over a fabric, one can allow oneself to be infused with the experience of that fabric and yet not be giving that experience any form, i.e. be appreciating it to be perhaps so-and-so (e.g. rough or smooth) or making any justifiable or systematisable claim about it being so-and-so (e.g. rough because it has such-and-such bumps, or rougher than the previous fabric because it scrapes the skin more). It is, put differently, a thin kind of experience – and yet still active and responsive (it is, for instance, quite unlike sleepwalking or daydreaming).

In summary, then, the second-person perspective shows us that social life ought to be understood as a mix or balance of the regular and the irregular, where, in addition, persons one interacts with are always to some extent experienced as other in a way that is neither immediately, nor perhaps ultimately, understandable. For persons

to be able to participate in social life, conceived of in this way, they must have abilities that allow them to be, to some extent, hesitant and tentative in their relations with others, and thus tolerant of ambiguity, uncertainty and unpredictability, and responsive to and capable of learning from the otherness of others. These abilities – or, differently put, cognition understood from the second-person perspective – can be profitably described as a partnership of appreciation and attentive silence. Persons must be able to attend silently to others, going on, potentially to appreciate them as *perhaps* being so-and-so or *perhaps* doing something in such-and-such a way. This can include appreciating something someone does as *perhaps* appropriate or inappropriate, with such-and-such actions of one's own in response being *perhaps* appropriate or inappropriate. The point is that these abilities – attentive silence and appreciation – are both tentative and hesitant, and, as a result, they allow us to picture persons as, at least in part, active and creative vis-à-vis that which can emerge as relevant, including normatively relevant, in the course of interaction. It is, then, these abilities that Khushim was missing, and his not having them is the reason – or at least a possible reason that should not, by definition, be excluded – for him not being able to get along with others.

Given this approach to the social, and our ability to participate in it, it is imperative that we do not construe the normative dimension of social life in such a way that it excludes what the second-person perspective allows us to see. This is important not only for the sake of not misunderstanding the social, or for the sake of adequately describing the abilities that we need to participate in it – it is also for the sake of including, within any account of the normative dimension of social life, those tentative and hesitant experiences that persons have about what they ought to do or about what might perhaps be appropriate or inappropriate, i.e. precisely experiences of the kind that are identified by the second-person perspective.

Incorporating the second-person perspective does mean we have to make some changes to the way we think about the normative in general, and the normative dimension of social life in particular. It does not mean giving up on the distinction between the normative and the regular – that continues to be fundamentally important – but it does mean not excluding, as part of social life and as worthy of explanation, all that which is irregular. A radical way of putting it would be to say that there must be a sense in which the irregular is part of the normative. A less radical way, and the way adopted by this thesis, is to say that any account of the normative dimension of

social life must not be such as to exclude the importance of irregularity from social life. As we shall see, this will mean 1) not characterising conventions, norms and rules as determinants of appropriateness and inappropriateness; 2) not thinking of them as necessary; 3) not thinking of them as necessarily governing minds; and 4) not thinking of them as necessarily shared.

B. The Structure of the Thesis

The thesis is structured in six Parts, each containing between two and five Chapters, which are further broken down into Sections. Thematically speaking, the relevant global division is between Parts I to IV, and Parts V to VI. Parts I to IV articulate and criticise the governance view, and Parts V to VI articulate and illustrate the second-person perspective on social life and the abilities persons need to participate in it.

As was mentioned above, the governance view is broken down into four aspects, with the last three identified as 'elements' that are argued against, and the first as a characteristic. This first aspect, or characteristic, concerns the way the governance view understands conventions, norms and rules. The governance view sees these as determinants of appropriateness and inappropriateness. From the second-person perspective, they are but contingently useful normative resources. However, before the approaches of the governance view and the second-person perspective can be contrasted in this way, the thesis needs to provide an account of what conventions, norms and rules are. The first Part of the thesis performs both tasks, i.e. it first articulates how conventions (Chapter 1), norms (Chapter 2) and rules (Chapter 3) are often used in the literature, and then rationally reconstructs them (Chapter 4); and second, it contrasts the way in which the governance view and the second-person perspective approaches them (also Chapter 4). Chapter 4, then, insofar as it already contrasts the governance view and the second-person perspective, introduces some of the themes that arise in subsequent parts. That is inevitable, but hopefully also useful rather than confusing.

Part II then turns to the first element of the governance view, i.e. the claim that determinants of appropriateness and inappropriateness (as the governance view characterises conventions, norms and rules) are necessary. It articulates four versions of this first element: first, the argument from the very concept of appropriateness (Chapter 5); second, the argument from the possibility of error (Chapter 6); third, the constitutivist argument, or put differently, the argument from the nature of an activity

(which we shall examine in two different forms, one in terms of a theory about the nature of games, and the second as a theory that treats language as if it were a game) (Chapters 7 and 8); and fourth, the 'super norm' argument, or the argument that postulates rules or norms that are inherent but invisible, *a priori* and implicit (but potentially able to be made explicit) (Chapter 9). It is claimed, *contra* the governance view, that the very concept of appropriateness does not require determinants of it; that we can make sense of error without determinants of what is mistaken; and that the constitutivist and super norm arguments are both flawed, or at least have serious limitations to our understanding of the social, and our ability to participate in it.

Part III focuses on the second element of the governance view, namely the idea that the mind is necessarily governed by determinants of appropriateness and inappropriateness. It explores five ways in which theorists have sought to make plausible the way in which the mind is governed by those determinants: Chapter 10 focuses on the rule-following literature; Chapter 11 on the notion of computation, with a particular focus on the acquisition and exercise of linguistic ability; Chapter 12 on the idea that experience is necessarily mediated by concepts, which are conceived of as having a rule structure (the focus here is on the work of John McDowell); Chapter 13 on the idea of action 'being adequate, yet unreflective' (here, the recent work of Erik Rietveld is discussed, as it is brings together both classical readings in phenomenology and more recent contributions in cognitive science and neuroscience); and, finally, Chapter 14 on accounts of tacit knowledge, which present the mind as oriented by determinants of appropriateness and inappropriateness, though ones it does not know and possibly could never make explicit (here, the recent work of Harry Collins is analysed). Again, as in Part III, these approaches are examined carefully, but are on the whole resisted: all of them, it is argued, miss certain aspects of how the mind learns and works that are of vital importance to social life, and our ability to participate in it, as understood from the second-person perspective.

Part IV, entitled 'Necessarily Shared Governance', focuses, as the title suggests, on versions of the argument that determinants of appropriateness and inappropriateness are necessarily shared – this being the third and final element of the governance view. In fact, saying 'argument' here is a little misleading, for in many theories this is not explicitly argued for, but is simply assumed. The five Chapters focus on: in Chapter 15, on the notion of script / scenario following (in the work of

Christina Bichierri and David Velleman); in Chapter 16, on social conventions (primarily in the recent work of Andrei Marmor); in Chapter 17, on one version of the appeal to a social order, i.e. via the notion of practices (with special attention to the work of Joseph Rouse); in Chapter 18, on another such version, this time by reference to the notion of games (with particular focus on Erving Goffman's work); and Chapter 19, on a number of different versions (including those of Pierre Bourdieu and Hans Joas, as well as some recent social cognitive theories) of the claim that we act unreflectively, yet in socially appropriate or adequate ways. Similarly, here, the thesis argues against those views insofar as they render it impossible to see those aspects of social life that the second-person perspective brings into view, and thus also makes it impossible to see how persons could participate in social life so conceived.

As noted above, Part V marks a shift in the thesis, both in terms of the material covered, but also in terms of the tone: the focus here is on positively building up a picture, rather than unravelling one it is argued we need to resist.

Part V outlines what is meant in the thesis by a 'second-person perspective'. It does so in four Chapters: Chapter 20 looks at the work of Vasudevi Reddy on infants' second-personal understanding of other minds (this also allows us to mark an important distinction between a certain understanding of games and the notion of play); Chapter 21 consider Hanneke De Jaegher's notion of the irregular rhythm of social interaction; Chapter 22 examines the specific notion of interactional breakdown from the second-person perspective, i.e. where breaking down is the result of too much, and not too little, regularity, order and patterning; and Chapter 23 on the specific sense of the otherness of others, or difference, that is fostered by the second-person perspective, namely not an immediately identifiable and intelligible difference, but precisely a difference that is unfinished and incomplete (thereby enabling the experience of the other as an autonomous source of invitations or opportunities for relating with them).

Having outlined what the second-person perspective reveals about social life, including interaction going better or worse, Part IV turns to describing the abilities that persons need in order to participate in social life, including having certain normative experiences in social life so conceived. This Part does so in two Chapters: in Chapter 24 it describes the ability to appreciate, and also distinguishes the context of appreciation from the contexts of justification and systematisation; and in Chapter 25, it articulates the ability of attending silently to others, including showing, in the

last Section of that Chapter, how appreciation and attentive silence work in partnership to enable persons to participate in social life, when it is understood from the second-person perspective. This Part, then, illustrates cognition – or cognitive abilities – from the second-person perspective.

Before moving on to outlining the sources and scope of the thesis in the next Section, a note should here be given as to the above-mentioned difference in tone between Parts I to IV and Parts V and VI. It is true that the tone in the first four Parts is negative. However, every effort has been made to engage carefully and generously with the work of the theorists that are considered. This renders the thesis much longer than it might have otherwise been. Given, however, that criticisms are being made, the thesis adopted an approach that takes its time, conversing with the theorist's work slowly and patiently, trying to understand how it comes to the views it espouses, what the benefits and importance of those views are, and how those views are limited in terms of what the second-person perspective shows us we need to take seriously about social life and our participation in it. If the thesis is at any point unfair to the theorists it discusses, it is not for want of trying. Further, it should be remembered that the second-person perspective by no means counsels against the importance and usefulness of some – indeed many – aspects of those views that are analysed here under the canopy of the governance view; on the contrary, it attempts to soften that view, and thereby in fact to show that many aspects of it, when properly acknowledged to be limited, are of great and lasting importance to our understanding of social life and our abilities to participate in it.

C. Sources and Scope

There are a number of matters of sources and scope that need to be mentioned in this Introduction. The first concerns the use of the term 'second-person'. This term has appeared in various literatures, or something similar to it has been developed though under different terms, and this thesis needs to clarify where it is positioned vis-à-vis those previous uses. The second concerns the use of the term 'normative', including the distinction sometimes made between the normative and the evaluative, and also the debate between the so-called 'normativists' and the 'naturalists'. The third is related to the second and concerns the variety of different kinds of conventions, norms or rules, and whether this thesis, when it refers to conventions, norms and rules

in general, wishes to refer to all those different kinds. Let us, then, proceed in that order.

The Second-Person

Reference to the second-person, and related ideas, has occurred in a great many contexts. It has, for instance, appeared in 1) the literature on social cognition; 2) the literature on the methodology of social science and the philosophy of history; 3) the moral philosophical literature; 4) linguistics and communication studies; and 5) literary theory. Even that list is no doubt not comprehensive. Given these various uses – which we shall discuss in a little more detail in a moment – this thesis does not claim to be *about* the idea of the second-person. It is certainly not a history of the idea, though telling that history would be an exciting project. There is also little doubt that the second-person perspective developed here could benefit enormously from a careful and detailed reading of the above-mentioned contexts in which reference has been made to the second-person, or related notions. Again, such a task falls, regrettably, outside the scope of the thesis, though, once more, it would constitute a logical and promising sequel to this thesis. Having made these preliminary observations, let us briefly consider the above contexts, in the order listed above.

Reference to the second-person has sometimes occurred as part of an attack on prioritising the 'third-person stance', or the 'observational', 'theoretical', 'spectatorial' stance (see, e.g. Hutto 2004 and Gallagher 2009) in theories of social cognition. Daniel Hutto, for instance, calls for abandoning 'the spectatorial stance' on the basis that it is overly theoretical, requiring persons to make inferences as to the allegedly hidden mental states of others 'for the purpose of prediction, explanation and control' (Hutto 2004, 549). Similarly, Shaun Gallagher's Interaction Theory (IT), which is designed to be an alternative to Simulation Theory (ST) and Theory Theory (TT) approaches to social cognition, is based on the idea that ST and TT 1) 'frame the problem in terms of the lack of access that we have to the other person's mental states'; 2) assert that 'our normal everyday stance toward the other person is a third-person observational stance'; and 3) hold that 'these mentalising processes constitute our primary and pervasive ways of understanding others' (Gallagher 2009, 291). Gallagher's IT theory is designed primarily to avoid the kind of commitment to inferential postulating of others' mental states that he sees as connected to the analysis of interaction from the third-person stance.

These are important criticisms of the third-person stance, but both Hutto and Gallagher locate the problem with that stance as being about the mode in which one cognises others, i.e. more precisely, whether it involves making inferences. As a result of locating the problem there, they both 1) do not defeat this way of thinking about how we relate with others (for, as many have pointed out, the inferential processes can simply be said to be sub-personal and not phenomenological);[2] and 2) they end up still positing allegedly shared structures that, according to them, drive social cognition. In terms of the second point, Hutto does this by arguing (on occasion in conjunction with Gallagher, see Gallagher and Hutto 2008) that we understand others via narratives, i.e. we have shared expectations, sourced in common narratives (as a result of 'hearing and learning about narratives'), that 'underwrite our ease in understanding one another' (Hutto 2004, 550). Gallagher, in turn, makes this second move by saying that the 'direct perception' we have of others (he says: 'we have a direct perceptual grasp of the other person's intentions, feelings, etc': Gallagher 2008a, 535) is 'already informed by my own interactions with her and others, as well as by my previous situated experiences, by my habitual ways of understanding, by socially established practices and by narratives that carry with them cultural norms' (Gallagher 2008c, 168). Gallagher also speaks of that direct perception being situated in 'projects' or 'pre-defined relations' (Gallagher 2009, 292), or in 'pragmatic contexts' which 'frame our encounters with others' (Gallagher 2008b, 173).

The effect of postulating these shared structures is that Hutto and Gallagher squeeze out or minimise the importance of what is called in this thesis the second-person perspective. Gallagher, for instance, draws on the views of Vasudevi Reddy (whom we consider in Part V of this thesis) in developing his notion of 'primary intersubjectivity' (see, e.g. Gallagher 2009, 293). By this he means the 'basic sensory-motor capacities that motivate a complex interaction between the child and others' (Gallagher 2009, 292). Some of what Gallagher says about this primary intersubjectivity has affinities with what is here referred as the second-person perspective, but it is noteworthy that Gallagher establishes a hierarchy of 'intersubjectivities'. The primary intersubjectivity is said to be very primitive, and not itself sufficient for 'complex interactions' (which involve the understandings of

---

[2] See e.g. Herschbach 2008; the point here is not to endorse Herschbach's criticism, but to show how Hutto's and Gallagher's critique of the third-person stance loses bite when it is relegated to a 'phenomenological layer'.

'motives and reasons'; Gallagher 2009, 292). Although it persists to some extent, it is relatively quickly – beginning at year one – taken over by secondary intersubjectivity, which is 'based on the development of joint attention, motivates contextual engagement, and acting with others' (Gallagher 2009, 292). Finally, between the ages of two and four, we develop 'narrative competency',[3] and it is only thanks to this competency that we can engage in those complex interactions. The idea, briefly put, is that we acquire narratives, which 'carry with them cultural norms', such that when we find ourselves in (what we somehow recognise to be) such-and-such an environment we perceive persons as characters in those narratives and enact the narratives within which those characters play certain roles (thereby, on this view, behaving in socially appropriate ways).[4] This, in effect, is a version of the governance view[5] – social or cultural life is imagined to be governed by certain norms, and persons are said to be able to participate in that society or culture because they acquire a mode (via their narrative competency) of detecting when they apply and following them.

Very recently, and still in the context of the literature on social cognition, Michael Pauen has referred to 'the second-person perspective' (in a paper with that title: see Pauen 2012). By this he means, however, 'drawing on one's own experiences in order to ascribe mental states to another sentient being' (Pauen 2012, 34). This, in effect, is very similar to a simulationist approach to social cognition. This approach is certainly one that can tell us a great deal about how we might sometimes relate to others, and nothing said in this thesis is a direct argument against it (nor indeed the other theories of social cognition). But, for our purposes, it is of limited value, for it reduces the second-person perspective to having another person as an object upon which one projects one's own experiences in order to make sense of that other person. In that sense, it precisely says nothing about how persons are able to relate to, and be responsive to and learn from, the otherness of others – that vital

---

[3] Hutto, too, emphasises this narrative competency: see Hutto 2008.

[4] Gallagher is greatly influenced here by the Heideggarian tradition of phenomenology according to which we always and already perceive the world – or the world always appears to us – in some meaningful way, as 'encoded in terms of possibilities for action' (quoting Wheeler 2005 with approval: Gallagher 2008b, 176). This approach is resisted in this thesis precisely because the second-person perspective emphasises that persons need not find the world, or others, immediately and easily meaningful; instead, the hesitation and tentativeness that we experience in relating with others is very much of what makes interaction with others meaningful.

[5] More recently, Gallagher has collaborated with De Jaegher and Di Paolo 2010, and embraced what they call 'strong interactivity', but this view is still in early stages, and it is not clear how strong interactivity is related to (for instance, does it replace?) Gallagher's division of labour between primary and secondary intersubjectivity.

presence of difference that is one of the key features of the second-person perspective, as that term is referred to in this thesis.

Pauen's paper illustrates another problem with theories of social cognition from the perspective of this thesis. Their explanatory object or target is often characterised as a matter of explaining how persons understand other minds – how they are able to have access to the minds of others. From the perspective of this thesis, that is too quick. It is too quick because when we interact with another person we are not necessarily aiming to understand them – our interactions are often more pragmatic, and thus also less spectatorial or theoretical, than that; they are less about understanding, and more about responding or reacting. To say, in answer to this, that responding or reacting can only follow understanding is to beg the question: the point is that we do not need to understand others (in the sense of having access to the contents of another's mind) in order to interact with them. Not needing to have such access does not mean we are not engaging with others, in all their difference; in fact, to the contrary, in order to engage with others, in all their difference, is not to think that their mind is like a bucket with a certain content in it that we, by applying certain rules, can understand. Of course, we can approach others as if they are just like that – buckets to be looked into, or puzzles to be solved, or characters in a play whose next move has to be worked out – but then we are not interacting with others in the way that the second-person perspective can show us we can, and often do, interact with them.

Another context in which reference to the second-person perspective has been made is in the context of the methodology of the social sciences, and some related disciplines, such as the philosophy of history. James Bohman has, for instance, referred to 'second-person interpretation', but by this he means to refer to an ethical or moral dimension of the practice of ethnography (Bohman 2000). Thus, he says, that second-person interpretation is a matter of adopting a certain kind of 'normative attitude', i.e. one that 'incurs a commitment or obligation by interpreting what others are doing, by interpreting what they say as correct or true for themselves as well as those whom they are interpreting' (Bohman 2000, 227). In effect, this is to give the benefit of the doubt, as to the correctness or truth of certain statements made by those one, as the ethnographer, is studying. This is very different to the second-person perspective adopted here for, although the primary sense of the second-person perspective is a theoretical looking-glass – a way of seeing features of social life as

part of it, and as worthy of explanation and description that we might otherwise miss – it is not a strategy for interpreting the behaviour of others. The second-person perspective is, in that sense, also not a method for more faithfully reporting the views of actors in the past – as it has been recently referred to by Karim Dharamsi in the context of R.G. Collingwood's well-known theory of re-enactment (see Dharamsi 2011).

In the moral philosophical literature, the most prominent reference to the second-person appears in the work of Stephen Darwall. The culmination of this work was Darwall's book, *The Second-Person Viewpoint* (2006). For Darwall, as he summarised it recently, second-person relations are dependent on 'authority and accountability relations': first, there is 'the authority to make a claim on or demand or expect something of someone'; second, there is 'an authoritative (legitimate) claim or demand'; third, there is 'a (second-personal) reason (for complying); and fourth, there is the fact of 'being accountable (to someone with the requisite authority) for complying' (Darwall 2010, 217-8). Understood this way, the appeal to the second-person offers a way of understanding how the role of rules that impose moral obligations changes when they are made in the form of a demand by someone directed at us. For example, Darwall says: 'When…we hold someone, ourselves or someone else, accountable for violating a moral obligation, and blame her for doing so without adequate excuse, we make a demand on her conduct not as individuals, but as representatives of the moral community… we give voice to demands that could come from any of us and that we presume to come from all of us' (Darwall 2010, 223). As we can see in that passage, in Darwall, the force of the third-person is always behind the second-person demand; the second-person demand is a way of expressing or putting in action third-person requirements (those rules that govern that moral community). Darwall's contribution is important for understanding one of the ways in which moral obligations can confront us, but in that respect, his use of the term 'second-person viewpoint' is much too narrow for our purposes. Further, insofar as it presupposes the pre-existence of a third-personal source of rules that give content to the second-personally expressed demands, it is somewhat at odds with the spirit of the second-person perspective offered in this thesis. That is because the second-person perspective, as it is used in this thesis, is designed to show us precisely the inadequacy of an approach to the social that begins with – that presupposes the existence of – certain rules or requirements. To begin in that manner is to presuppose an order – in

Darwall's case, a moral order – that then constitutes the framework or structure within which one describes how persons manage to relate to each other (again, in his case, make moral demands of and behave morally towards each other). That beginning is much too thick for the purposes of this thesis. Of course, this is not to criticise Darwall – it is just to point out that his use of the second-person is quite different to the one adopted here.

A fourth context in which the second-person has been appealed to is in the study of language and communication. Thus, for instance, Donald Davidson (1992) has appealed to the second-person because he wants to make more central, in our inquiries about the nature of language, the scene of interaction between a speaker and an interpreter, and thus to re-orient focus away from 'the abstract character of language' (Davidson 1992, 256) to the pragmatics of communication.[6] This kind of re-orientation is broadly within the spirit of this thesis, but Davidson's specific treatment still refers to the rather peculiar relation of an interpreter to a speaker, i.e. a relation in which one person is the passive object of the other's active cognitive exercise. This is a kind of inter-personal relation that very quickly turns into an oscillation between first-person anxiety concerning the status of one's interpretations of the other and third-person appeal to 'the public character' of language (Davidson 1992, 262). It is, in essence, not interactive enough – it does not make enough room for a person to be responsive to and learn from the otherness of the other, and thus makes it difficult to see how persons can create novel ways of relating with that other in the course of the interaction.

Other approaches to the study of language and communication – such as the 'dialogical' approach – are also relevant to mention here. The idea that meaning is constructed in the course of an interaction conceived dialogically (see, for instance, Grossen 2010) is certainly a promising future source for further elaborating on the second-person perspective in the future. Many of the dialogical approaches to linguistics are inspired by, at least in part, the work of Mikhail Bakhtin, whose reading of Dostoyevsky (see Bakhtin 1984 [1963]) has of course been greatly influential in many scholarly contexts,[7] including the fifth context we wished briefly

---

[6] 'My present aim', says Davidson, 'is…to emphasise, following Grice, the central role of intention in communication' (Davidson 1992, 258).
[7] For example, Bakhtin's work influenced Vasudevi Reddy's work in developmental psychology, which we look at in Chapter 20 (see Reddy 2008, 34).

to mention, namely literary theory.[8] Bakhtin's approach is valuable, for, as Michèle Grossen notes, it pictures discourse as 'always addressed to an interlocutor (addressivity) and is therefore based upon an anticipation of the interlocutor's comprehension (responsivity)' (Grossen 2010, 7). Methodologically, studies of language and communication that follow Bakhtin's approach pay attention to 'an intersubjective process implying reciprocal attunement', including looking at how 'meaning is negotiated through interactional work' (see Grossen 2010, 8). These are studies that clearly take the presence of the otherness of the other seriously, and in that respect are close to the second-person perspective adopted here.[9] Further, to the extent that such studies incorporate ambiguity, uncertainty and unpredictability – in short, irregularity – into that 'interactive work' or that 'intersubjective process' or that 'negotiation', they are even closer to the way this thesis uses the second-person perspective.[10]

The literatures that refer to the 'dialogical' or to 'dialogics' is vast – they include, for instance 'dialogical phenomenology', as promoted by Barbara Stawarska (see Stawarska 2009), or 'dialogical game theory' (see Lecomte and Quatrini 2011). As a group of related ideas, the dialogical literature might also find resonance in 'interactionist' approaches to cognitive science (e.g. Bickhard 2009) and developmental psychology (e.g. Rochat 2009; Boden 2006).

There are, of course, many other theorists who might be thought to have emphasised the importance of dialogue and interaction, in both social science and social psychology. The reason for pointing to such links is just to give an indication of other literatures that may be sympathetic to the direction taken in this thesis, and that could also be drawn on in the further development and perhaps generalisation of the value of the second-person perspective. Delving into them in any more detail is outside the scope of this thesis.

---

[8] See also Rifelj (1992), who offers an exploration of the other minds problem in a variety of literary works.

[9] See, further, Marková 2003.

[10] Related references appear in Grossen (2010). See, for instance, the work of Ragnar Rommetveit (see Rommetveit 2003 and Linell 2003). Per Linell has recently offered a whole-scale overview of this approach: see Linell 2009. Murakami (2010) in a response to Grossen (2010) notes the work of Victor Turner on liminality (see Turner 1969) and says it can help understand 'the emergent, in-between and temporal nature of intersubjectivity in interaction' (2010, 32). The recognition of asymmetrical subjectivities engaging in an interaction that can transform them both in ways that neither could predict, or be prepared for, can be an important resource for further developing the second-person perspective.

The Normative, the Evaluative and the Normativist-Naturalist Debate

When reference is made in this thesis to 'the normative', it is done so in the widest possible sense. It thus does not subscribe to the terminology that is used when a distinction is made – and usefully so in many contexts – between the normative (then used in the narrow sense) and the evaluative.[11] Thus, for example, John Broome, for whom 'ought is the central normative concept', says: 'Many authors include the evaluative within the normative. I do not. So for me good is not a normative concept at all. I am neither asserting nor denying that ought is in some way more central than good' (Broome unpublished draft, 8). This attempts to be neutral with respect to the evaluative (and whether it is reducible to the normative in the narrow sense), but it does – as a result of adopting the very distinction between the evaluative and the normative – greatly narrow the ways in which we can speak of the 'normative'. A similar move is made by John Skorupski, for whom the concept of a reason (rather than ought) is 'the fundamental normative concept' – though fundamental in the sense that 'it supplies illuminating terms for…the distinction between the descriptive and the normative' (Skorupski 2007, 247), and not the distinction between the normative and the evaluative (see Skorupski 2007, 249). Again, this remains neutral as to the evaluative, but it narrows our use of the term 'normative', and makes us think of it in the narrow sense (of reasons or oughts).[12]

At other times in the literature, 'the priority of the normative' is referred to precisely in order to argue for the reducibility of the evaluative to the normative (in the narrow sense).[13] The phrase, 'the priority of the normative', is used in a recent

---

[11] Truth be told, matters can get very complicated very quickly here, as when there are two sets of oppositions in the air: 1) the normative versus the non-normative (where, occasionally, the non-normative will be expressed in other ways, e.g. the 'descriptive'); 2) the normative versus the evaluative (not to mention possibly a third: the evaluative versus the non-evaluative). Sometimes, in contrasting the normative and the non-normative, theorists will employ concepts (for instance, they refer to 'reasons' and 'oughts') in ways that it is unclear whether or not they intend to assert that the evaluative can be reduced to reasons or oughts (and thus to the normative in the narrow sense). But even when a theorist makes it clear that they intend to be neutral on the issue of whether the evaluative is reducible to the normative in the narrow sense, their very use of the normative as independent from the evaluative narrows the usage of the term 'normative'.

[12] It should here be added that there are many other terms used aside from the normative versus the evaluative to mark the same, or a similar distinction, e.g. Judith Jarvis Thomson (2008) refers to 'evaluatives' versus 'directives', and Stephen Finlay speaks of 'mattering' versus 'favouring (see Finlay 2006 – for a general overview, see Finlay 2010). Of course, as similar as these are, no claim is here made that they are the same, e.g. to refer to evaluatives and directives suggests a reference to propositional statements, whereas referring to mattering and favouring refers to attitudes.

[13] When this is so, no explicit claim need be made as to the best way to distinguish between the normative and the non-normative (though it is surely implied that if all values can be reduced to, e.g. reasons, then reasons are also the distinguishing mark of the normative vis-à-vis the non-normative).

paper by Jennie Louise (2009), who also points out that so-called the 'buck-passing' or 'fitting attitude' analyses of value have come to dominate the domain of meta-ethics. A lot can be claimed for reasons in this vein, e.g. consider Joseph Raz's claim that 'the normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons… the explanation of normativity is the explanation of what it is to be a reason, and of related puzzles about reasons' (Raz 1999, 354-5; where reason itself is 'inherently normative' for Raz: 1999, 355). An often-mentioned motivation for this prominent placing of normativity in the narrow sense of 'reasons' or 'oughts' is that it avoids postulating queer properties. This is claimed because it is arguably easier to agree on the existence of reasons or oughts than it is on other kinds of properties proposed by metaphysicians of value.

In all these cases, then, the normative is used in a narrow sense to refer to reasons or oughts, which are then typically said to the basic normative concept – in some versions, so basic that even the evaluative can be reduced to one of them. For the purposes of this thesis, we need not enter this quarrel. What is of more importance to us is avoiding, at the outset of our investigation, the narrowing of the understanding of the normative. The reason for avoiding this narrowing is that it has certain effects on what is then identified as relevant to the analysis, especially in terms of the eventual descriptions of the abilities of persons to participate in social life.

We can get a sense of that effect when we consider the recent collection by Simon Robertson (2009a) who, in his 'Introduction', says that 'normative is that which constitutes, serves, implies, is derived from or prescribes some norm', and where any such norm is itself said to be profitably understood in terms of reasons or oughts (Robertson 2009b, 1-2). The relevant point, then, for present purposes is that although it need not be so[14] talk of reasons or oughts tends to sit better with talk of norms (or conventions or rules, all characterised as determinants of appropriateness and inappropriateness).

That the normative in the narrow sense (of reasons and oughts) sits well together with talk of norms and rules has been recently helpfully pointed out by Peter Railton in a forthcoming paper on Nietzsche and normativity (Railton forthcoming).[15] By 'normative concepts proper' Railton means to include the following: 'rule, norm, standard, law, right, wrong, correct, incorrect, rectify, regulate, require, permit,

---

[14] Jonathon Dancy's particularism about reasons may be an exception, see Dancy 2004.
[15] Reference here is made to a draft kindly emailed by private communication.

prohibit, duty, obligation', and in the category of the evaluative, he lists 'good, bad, noble, base, fine, magnificent, great, desirable, rewarding, virtue, virtuous, vice, vicious, worthy, worthless, admirable', etc (Railton forthcoming, 5). There are four characteristics, he says, of normative concepts proper: first, in terms of fit, they focus on 'formal correctness' by way of conformity to a standard; second, they place emphasis on the voluntary, both in the sense that 'ought implies can', but also more broadly in that these concepts are associated with acts of which compliance can be required; third, these concepts are discrete, in that e.g. norms and rules are typically binary and discontinuous, and function in the guidance of action or deliberation as conditions that must be met; and fourth, they are concepts that are exclusionary, in that conflict between rules or duties requires adjudication to reach practical conclusions about what ought to be done. By contrast, evaluative concepts are characterised by: first, a fit that focuses on a substantive match, e.g. realising or harmonising with a nature or *telos*, or meeting a need, or completing or perfecting a whole; second, they are concepts that are generally non-voluntary in nature, e.g. the phenomenon of respect; third, they are continuous, so that they can be realised to a lesser or greater degree; and fourth, they are non-exclusionary, meaning that conflicting values can not only co-exist but sometimes also be promoted in a single act.

Railton's list of distinguishing characteristics is instructive (though it should be added that it is not being endorsed here – it is just being referred to in order to make a different point), for it shows how talk of reasons and / or oughts sits well with talk of rules and norms. If we place our bets, then, on marking the normative by reference to the normative in the narrow sense (leaving aside, even if not reducing, the evaluative) then we may be prone to emphasising the role of determinants of appropriateness and inappropriateness, for those determinants are devices suited for voluntary guidance or deliberation about what I or we have reason to or ought to do. When we do so, however, we may miss important ways in which the evaluative (or the normative in the broad sense) might make us position the agent differently. Indeed, we see this in Railton's distinction, for he points to the non-voluntary, non-exclusionary, and continuous nature of the evaluative, which, very importantly, then trickles down to produce a different image of normative cognition.

Again, the purpose of making these remarks is not to agree or disagree with claims concerning the relationship between the normative and the evaluative. Rather,

it is to notice, and to raise concerns about, the effects – particularly on our picture of cognition, including recognising the kind of normative experiences we can have – when we place all our bets on the normative in the narrow sense. For these reasons, this thesis does not adopt the terminology of the normative versus the evaluative. The point it does make is that a terminology that insists on a strict division between the normative and the evaluative is one that is more likely to appeal to the governance view – again, the reason for the appeal is that talk of reasons and oughts in the narrow sense sits well with the explanatory emphasis given by the governance view to determinants of appropriateness and inappropriateness. This does not mean, of course, that this thesis does not recognise the importance of concepts like reasons or oughts, and thus the contributions of those theorists who have offered such illuminating insights about how such terms are used in daily life. All it means is that it resists the strategy that would adopt that narrow reading of the normative from the beginning of the inquiry. Ultimately, of course, it may be that talk of reasons or oughts is perfectly compatible with what the second-person perspective allows us to see about social life, including the kinds of non-convention, non-norm, and non-rule based normative experiences we can have when interacting with others – but that would arguably require a revision of how oughts and reasons are presently used in the literature on normativity. Furthermore, the cognition pointed to by the second-person perspective – that above-mentioned partnership between appreciation and attentive silence – is one that is at least *prima facie* more congenial to the evaluative side of any normative-evaluative split. Hence, again, the reason for not narrowing our understanding of the normative from the outset of the inquiry, including avoiding endorsing, for the purposes of this thesis, the distinction between the normative and the evaluative.

Apart from the normative-evaluative debate, the other debate in the literature concerns 'normativists' on the one hand, and 'naturalists' on the other hand (see, e.g. Turner 2010). A lot depends on how this debate is set up, and by no means does this thesis aim to make a contribution to it; indeed, to be properly understood, the debate would no doubt need to be situated in the context of the history of ideas, where, for instance, what might be called 'the demarcation strategy' (i.e. an approach to the study of human behaviour that tries to articulate what is allegedly distinctive about human life, as opposed to animal life) has been at play. Understood as connected to the demarcation strategy, the debate could be seen to be one where the demarcationists have sought to argue that what is distinctive about human beings is

that, unlike animals, they are governed by rules; animals, on this view, are instead governed by mere regularities. Put into the language of normativism and naturalism, the debate is then between those who argue that human life can only be studied normatively, i.e. by reference to the rules that govern human life and how humans rely on and use rules, and those who argue that any such rules, and their usage, are reducible to an explanation based on scientific laws.

The sympathies of this thesis clearly lie against the demarcationist strategy and, in that respect, also against normativism. However, in truth, the debate is orthogonal to the aims of the thesis. Perhaps the key point is that the problem with the debate is that both assume or posit an order, either composed of rules (for the normativists) or of scientific laws (for the naturalists). In so doing, the conception of the social is – or so it is argued in this thesis – watered down. As a result of such narrowing of the concept of the social, we miss what is revealed to us by the second-person perspective. Human life – or social life – is not be understood either by characterising it as governed by rules or scientific laws; and that is because it is not best understood as being *governed* – or being ordered, or patterned – at all. That way of talking about human beings or social life is misleading. This is because, as has been noted above, adopting the second-person perspective shows us that social life ought to be understood as a mix or balance between the regular and the irregular, the ordered and the disordered, the patterned and the un-patterned.

Taking this approach to that debate does not mean that the thesis is 'mystery-mongering'[16], as is often said of accounts that refuse to reduce the study of human behaviour to the operation of scientific laws; but neither does it mean that the account is anti-naturalist. One can, arguably, be a naturalist without committing to a view that human beings or social life are to be understood as being governed by scientific laws; for instance, naturalism could, instead, be understood as an attitude one takes to one's own theories (whether they be of animals or of humans) – an attitude of modesty and awareness of the defeasibility and limitations of any theory, of making sure a theory is testable and testing it.[17] Again, the above statements are made tentatively because it is clear that the debate between the normativists and the naturalists depends a great deal

---

[16] The charge of 'mystery-mongering', sometimes made by naturalists against normativists, is referred to, for instance, by McPherson 2011.

[17] In other words, one could take the same stance to one's theories as this thesis recommends we take to conventions, norms and rules, namely that they are contingently useful but can, in certain contexts, be experienced as necessary.

on how it is set up, how its terms are understood. To repeat, if that debate is understood as one between those who argue that social life is governed by rules and those who argue that, even if it is, the governance of those rules is reducible to an explanation on the basis of scientific laws, then that debate is orthogonal to the purposes of this thesis – and it is orthogonal because this thesis argues that the conception of the social, implicit in both sides of the debate, is mistaken. The question is not whether rules are reducible to regularities, but whether social life can be exhaustively understood as a matter of either rules or regularities. The point of the second-person perspective is to argue it cannot. That, in itself, to repeat once more, is not anti-naturalist.

The Variety of Conventions, Norms and Rules

This thesis recognises that there are many different kinds of conventions, norms and rules. It recognises, for instance, that, given the very different way these terms are used, conventions may sometimes be contrasted with norms on the basis that the former are 'merely conventional', likely to be culturally specific and not necessarily attracting moral criticism when deviated from, whereas the latter are 'moral', possibly culturally universal and certainly attracting moral criticism. The development of moral knowledge or moral judgement is sometimes said to be precisely the ability to be able to distinguish between conventions and norms understood in this sense (see, e.g. Turiel 1983, and more recently, partly in relation to the study of psychopathic behaviour, see Nichols 2004). In other words, when speaking of conventions, norms or rules one could be speaking in the context of a social activity that has, at least ostensibly, no moral overtones or moral implications (e.g. playing games, such as chess or snooker) or in the context of practices that clearly have such overtones or implications (for instance, any activity where someone's vulnerability or dignity is at stake, whether they be citizens vis-à-vis the government or children vis-à-vis their parents). There are, no doubt, other contexts still, such as aesthetic contexts or, perhaps even more specifically, the contexts in which logic is used or in which language is spoken.

This thesis recognises this context-dependency and acknowledges that, as a result, any general theory of the social, including in its normative dimension, will always be somewhat questionable, or at least capable of being confronted by the specificity of some particular context. Furthermore, it also recognises that no general

theory of conventions, norms or rules may be plausible or desirable outside certain specific contexts. However, be that as it may, it is nevertheless the stance of the thesis that whatever context one focuses on, the second-person perspective can reveal important features of the way human beings relate to each other that may not be revealed in any other way. The second-person perspective can still be valuable as a general looking-glass, even if its explanatory value will differ from context to context.

Having said that, the examples the thesis discusses, and the literature it focuses on, is most definitely not, or at least not centrally, moral. Occasionally, the thesis will revert to literature that is situated in the context of the use of logic, of reasoning more generally, or of the use of language. Where this is considered, it is done so for the purposes of examining the arguments made, insofar as they are extractable from that particular context. The focus, in short, is on the type of argument, e.g. the idea that there can be conventions, norms or rules that are necessary (for instance, *a priori* necessary) and the point being made is that insofar as we are discussing the social, including in its normative dimension, we are not talking about conventions, norms or rules that are necessary. Whether exceptions ought to be made for logic or morality or reasoning is another matter that falls outside the scope of the thesis. However, since the arguments made by theorists in these other contexts are sometimes appealed to or transposed from those contexts into the more general domain of our understanding of the social, including its normative dimension, so those arguments deserve a hearing.

Prima facie, then, when this thesis refers to conventions, norms or rules it means those conventions, norms or rules that are not somehow dependent on or particular to very specific contexts, like those of morality or logic. It intends to refer, instead, to those conventions, norms or rules that do not clearly have moral overtones or implications, i.e. for instance (and again, without wishing to delve into the quagmire of moral theory), where a person's very life or vulnerability or dignity is at stake.

Having, now, introduced the issues canvassed and arguments raised by this thesis, including having discussed the structure of the thesis, and certain matters of sources and scope, it is time to turn to the first Part of the thesis, namely a discussion and rational reconstruction of conventions, norms and rules.

Part I. Conventions, Norms and Rules

As noted in the Introduction, at stake is in this thesis is the conception of the social, as well as our ability to participate in it. If, from the beginning, we conceive of the social broadly, to include that which becomes visible from the second-person perspective, then we also improve our description of the abilities we need to participate in social life. If, however, we conceive of it narrowly at the outset, it is likely we will not only miss important features of social life, but we are also going to be less nuanced and subtle than we need to be in our description of the abilities persons need to participate in social life. One way of being too narrow in our conception of the social is by prioritising a certain view of its normative dimension. That view is the governance view, and this is the view that this thesis argues against. What the governance view does, essentially, is to impose, from the beginning, a narrow conception of the social in its normative guise. It does so by imposing, as an object of explanation, social order, which is created and maintained by a mixture of conventions, norms and rules.

As we shall see, on the rational reconstruction of these terms offered in this first part of the thesis, conventions and norms have a close affinity to that which is regular. They are not, however, merely regular; put differently, they are 'something more' than the regular. If they were merely regular, conventions and norms would arguably not be normative phenomena, given the fundamental distinction (especially for the governance view) between that which is normative and that which is regular. Conventions and norms have a close affinity to that which is regular because they require 1) the regular exercise of certain attitudes – these are attitudes as to what ought to be done; and 2) the regular exercise of that which is the object of those attitudes, i.e. precisely the repetition of that which (according to the attitude) ought to be done. We shall see, later in this Part, that conventions and norms differ in important ways, but for the purposes of this thesis, what they share, i.e. this connection with the regular, is what is critical.

By contrast, at least on the reconstruction offered here, rules do not have this connection to the regular; put succinctly, rules are 'something else', and not merely 'something more', than the regular. If one accepts the distinction between the regular and the normative then rules, compared to conventions and norms, are much more paradigmatic examples of the normative. Rules are neither attitudes, and they are not the object of those attitudes – nor are they a combination of the two. Instead, rules are

linguistically-expressed propositional entities issued by authorities that can feature – but need not (including not needing to do so regularly) – in attitudes about what we ought to do or as standards in the evaluation of ourselves or others. However, despite these differences with conventions and norms, rules are still very much connected to the governance view. For rules, like conventions and norms, are typically said to constitute a certain order, where, further, the internalisation of those rules is necessary to participating in that order. Where, then, rules become problematic from the perspective of this thesis is when 1) they play a part (either the dominant part, or in conjunction with conventions and norms) in constituting the object of explanation, i.e. they narrow, from the beginning, the conception of the social as a matter of an order; and 2) they then feature in the explanation itself, i.e. they are conceived of as being regularised in certain ways, e.g. through attitudes or skills that internalise rules, with the establishment and maintenance of order being explained precisely to be this internalisation (regularisation) of rules.

As we shall see, what this Part, and this thesis more generally, argues for is a limitation of the role of rules in the creation of the object of explanation (in this case, the social in general, but also the social in its normative dimension), and also for a correlated limitation of reference to rules in the explanation. The point, in short, is that the abilities we need to participate in social life are very much more than abilities concerning the internalisation or manipulation of rules. The catch is that we can only see we need those non-rule-based abilities when we enlarge the object of explanation, i.e. when we do not rely on rules (or, on conventions and norms) as setting up that which needs to be explained (about the social in general, or about the social in its normative guise).

It is certainly true that rules are vital resources we have, both for guiding ourselves (alone or in collaborative activity) and for evaluating ourselves and others. They are also highly flexible resources, capable of interpretation and re-interpretation and thus with dynamic applicability to new contexts. However, as useful as they are, they should not be permitted to define the object of explanation, nor dominate the explanation itself. By considering the social in general, and its normative dimension in particular, from the second-person perspective we can see that 1) the object of explanation should not be assumed or posited to be an order composed of conventions, norms and rules; and 2) that, therefore, the abilities we need to participate in social life are not only those that concern conventions, norms and rules.

Put more colourfully, the second-person perspective can help us see that there is more to the social than rules (or norms and conventions, for that matter) ever dreamed of, and thus more to the description of the abilities we need to participate in social life than any rule-based (or norm- / convention-based) abilities can secure.

In terms of conventions and norms, the thesis also seeks to limit their role both in terms of creating the object of explanation and in dominating the explanation – again, of the social in general, but also the normative dimension of the social in particular. In this sense, the argument against conventions and norms is akin to the argument against the regularisation of rules (the difference being that rules need not be regularised, whereas the very concept of conventions and norms is connected to the regular in the two ways noted above). Conventions and norms, then, narrow the object of inquiry because they impose, as the explanatory target, an order of behavioural regularities with regularly-exercised attitudes attached. What they miss, from the second-person perspective, is the irregularity in interaction. The point of the thesis is to argue that that irregularity is a vital part of the social, and a vital part of what we ought to be trying to explain.[18]

It is not, however, all bad news for conventions and norms as explanations. They are important, for they pick out certain kinds of attitudes we can have to what ought to be done – these being, precisely, attitudes that are exercised regularly and take as their object behaviour that is itself regularly performed. Being able to adopt certain attitudes – including those picked out by norms and conventions – are critical to persons' abilities to participate in social life. The point, for this thesis, is that there are more attitudes – and thus more abilities – that persons need than those picked out by conventions and norms. These are precisely attitudes and abilities that are not connected, in the way that conventions and norms are, to the regular. Hence, we should not permit the narrowing, from the beginning, of the object of explanation to the order or regularity established and maintained by conventions and norms.

---

[18] This can be a misleading way to put the argument, for the thesis does not want to depend on, or require, a dichotomy between order and disorder, regularity or irregularity, but the first step in the re-imagination of the social from the second-person perspective is adding, to the mix, disorder and irregularity, with future work to be dedicated to a new terminology, one that does not depend on this distinction between order and disorder or the regular and the irregular.

We shall be returning, in Chapter 4, to the points made in an introductory way in the above paragraphs.[19] But first, we need to try to get a handle on the way in which conventions, norms and rules (in that order) have been used in the literature and more pertinently for the purposes of this thesis, try to rationally reconstruct these uses. It is important to stress, at the outset, that the aim is not to be comprehensive in the review of the literature. The aim, instead, is to give a sense of the uses made, but thereafter focus on rationally reconstructing the terms.

---

[19] We shall also be discussing, in that Chapter, the issue of differentiating amongst different kinds of conventions, norms and rules, e.g. those in the domain of logic or morality on the one hand, and those concerned with social activities on the other hand.

Chapter 1. Conventions

There is an ambiguity, which ought to be mentioned at the outset, in the use of the term conventions. On some versions, the term 'conventional' is used in the same breath as the 'natural' or the 'regular'. Thus, when Wittgenstein says, in *The Blue and Brown Books*, that 'Here we strike rock bottom, that is we have come down to conventions' (Wittgenstein 1958, 24), he is referring to what is usually or commonly done – to 'how we usually do things here'. In a short paper on conventions, Hilary Putnam offers the following summary of Wittgenstein's view (also suggesting Quine held a similar one): 'Like Quine, Wittgenstein comes to the conclusion that the fact that in certain contexts we just "go on" the way we do – perhaps as the result of having a certain number of examples or of having watched members of our community or interacted with them in "language games" – is more fundamental and in every way prior to such activities as giving and interpreting explicit directions' (Putnam 1981, 4). As conceived in this passage, the conventional is said to be not only distinguishable, but also have a certain priority vis-à-vis, rules (at least rules understood as explicit directions). Putnam in fact defines the conventional not only in distinction to 'directions', but also in distinction to two other phenomena: the objective, as when we distinguish 'the entities which are really metaphysically there, and those that are the product of our invention' (Putnam 1981, 7; the latter, of course, are the 'conventional ones'); and the 'factual', meaning that 'everything we say is conventional in the sense that we might have said something else, perhaps something verbally incompatible; and everything we say is factual in the sense that we could not have said just anything else. The fact that we say X rather than Z may be conventional, wholly or in part, while the fact that we say X rather than Z is not at all conventional' (Putnam 1981, 10).

Such a use of the term 'convention' is not the central one referred to here. The central use is best examined by considering David Lewis's *Conventions* (1986), which has been a highly influential text in making the topic of conventions philosophically popular. Lewis's view is also useful to consider because he takes particular (and rare) care to distinguish conventions from norms and rules.

A. Introducing Lewis

Lewis's view is inspired, in broad terms, by David Hume, though exactly what Hume's view of conventions was is controversial. John Latsis, for instance, in a recent paper (2005), argues that there are two views of conventions in Hume. The minimum that Latsis finds is common to both of Hume's views is that convention is a solution to the problem of selfishness, though not one that requires an explicit agreement or promise; there is a sense in both, says Latsis, 'of common interest that generates rules' (Latsis 2005, 711; why the interest in question has to generate 'rules' is not explained).

In the first version, the focus is on conventions that 'come about in situations where a number of agents each stand to gain from the institution or perpetuation of a particular social practice amongst them. Thus, each individual gains by adopting a principle that regulates his interaction with his fellows, provided that they reciprocate' (Latsis 2005, 711). Apart from there being a coincidence of mutual interest, which can be achieved by the adoption of a common course of action (Latsis refers either to the adoption of a rule or to a principle) that is engaged in provided that (and presumably because) others do the same, 'each person must express (by some external signal) his or her knowledge about the mutual interdependence' and 'must also be aware that every other party has also understood that mutual interdependence obtains' (Latsis 2005, 711).

The second version, by contrast, 'abandons the clause about declaring mutual interests' as well as 'talk of principles', thereby becoming 'independent of linguistic practice' and capable of arising 'without signalling or the declaration of mutual interests' (Latsis 2005, 712), as when two rowers row in tandem without instructions. In the second version, conventions 'arise out of a type of inductive, trial-and-error based reasoning where language is not essential' (Latsis 2005, 712).

Lewis certainly takes from Hume the focus on mutual interdependence, or the coincidence of mutual wants / interests. He positions agents in so-called co-ordination problems where the task faced by agents 'is to co-ordinate around one among a number of equally satisfactory possible outcomes, given that any coordinated behaviour is better than none at all' (Latsis 2005, 712-3). Agents need to reach one of a number of possible 'equilibrium' states. One obvious way in which they can do so is where one of the possibilities is one that is well-known to the relevant population of agents, for it has occurred before. Here, typically, agents will not have any reason to

look for an alternative, and precedent will make the previously reached equilibrium the salient one. All this, says Latsis, 'leads to conformity in the actions of individuals, which is represented by a behavioural regularity that conforms to one of the possible equilibrium strategies of the coordination game' (Latsis 2005, 713).

Latsis himself finds fault with Lewis's account, mainly because even with the device of precedent and natural salience it is still too demanding cognitively, e.g. as Latsis characterises it, 'the only way for agents to reach an equilibrium is through calculations based on their expectations of what others think. This leads to a complex regress of higher order expectations. Agents can consult their higher order expectations by replicating the practical reasoning of others' (Latsis 2005, 713). Latsis thinks that Lewis's view comes closer to the first of the above two versions of Hume's convention, and that many contemporary views of conventions have been of this kind, much to the detriment of our understanding of conventions.

For present purposes, the evaluation of either of Hume's, Lewis's or Latsis's views is not at issue. What is of interest is to see how the term convention comes to be used, and where its use either identifies with or clashes with the notion of a 'mere regularity', and if it clashes, then in what way it does so.[20]

B. The Normativity of Conventions

To what extent Lewis thought conventions were normative – and thus, given the centrality of the distinction between the normative and the regular, 'something more' or 'something else' than the allegedly merely regular – is somewhat controversial. In his book on *Conventions*, Lewis first suggests that conventions simpliciter are distinguishable from conventions that may be considered to be 'a species of norms' (Lewis 1986, 97). However, although Lewis does distinguish conventions from norms, he also argues that in most cases it will be likely that conventions will be norms. Thus, he argues that 'there are certain probable consequences implied by the fact that an action would conform to a convention (whatever the action and whatever

---

[20] Lewis, it must be said, changed his own position and his definition of conventions (a fact, as Olivier Favereau 2008 notes, that is often overlooked). As reconstructed by Favereau 2008 taking into account Lewis's later revisions, Lewis's definition of conventions is one that places agents in situations where what matters most is that everyone chooses the same action or belief R (not so much what specific action or belief is chosen), such that: 1) everyone conforms to R; 2) everyone believes that the others conform to R; 3) this belief that the others conform to R gives everyone a good and decisive reason to conform to R himself; 4) there is a general preference for general conformity to R rather than slightly-less-than-general conformity; 5) R is not the only possible regularity meeting the last two conditions; and 6) all the above conditions 1 to 5 are matters of common knowledge (see Favereau 2008, 119).

the convention) which are presumptive reasons, according to our common opinions, why that action ought to be done' (Lewis 1986, 97; it being the presence of these 'presumptive reasons' that turns conventions simpliciter into conventions as a species of norms). The presumptive reasons in question are: first, that 'I have reason to believe that my conforming would answer to my own preferences', and second, that 'I have reason to believe that my conforming would answer to the preferences of most other members of P involved with me in S; and that they have reason to expect me to conform' (Lewis 1986, 98). Lewis follows this list of the presumptive reasons with the following statement:

> For we do presume, other things being equal, that one ought to do what answers to his own preferences. And we presume, other things being equal, that one ought to do what answers to others' preferences, especially when they may reasonably expect one to do so. For any action conforming to any convention, then, we would recognise these two (probable and presumptive) reasons why it ought to be done. We would not, so far as I can tell, recognise any similarly general reasons why it ought not to be done. This is what I mean by calling conventions a species of norms. (Lewis 1986, 98)

As is clear from the above statement, a norm dictates what ought to be done. A convention on its own does not do so, but given certain likely circumstances, it often will be a norm (and therefore also determine, at least by default, what is appropriate and what is inappropriate). Lewis goes on to show that where a convention is a norm, it is also a 'socially enforced norm: one is expected to conform, and failure to conform tends to evoke unfavourable responses from others' (Lewis 1986, 99). Analysing Lewis's sense of the social enforceability of convention as a kind of norm is not necessary for present purposes. What is important to see is how Lewis comes to add a normative dimension to conventions, i.e. by reference to a convention potentially (and usually) being a norm, and a norm in the sense of something that dictates what ought to be done.

Conventions, then, are in principle distinguishable from norms, though in practice are likely to be a kind of norm. The connection between conventions and rules, on the other hand, is not as close as the connection between conventions and norms (although, as we shall see in a moment, the usage does overlap to some extent).

In discussing the distinction between conventions and rules, Lewis does acknowledge that 'we would certainly call many conventions rules', especially 'tacit,

informal or unwritten rules' (Lewis 1986, 100), but he adds that there are many phenomena called rules – such as 'mere generalisations, laws of nature, or even mathematical truths' – that we would not (given his definition) classify as conventions (Lewis 1986, 100). This is largely because such rules 'may have nothing to do with the conduct of human agents, except that human agents might benefit by taking account of them' (Lewis 1986, 100). Exempt from status as conventions are also rules as 'strategic maxims', i.e. 'hypothetical imperatives stating what a human agent might do to gain some end. These rules state generalisations regarding the tendency of certain actions to accomplish certain ends' (Lewis 1986, 101). Excluded also are rules that 'are hypothetical imperatives reinforced by authoritative codification and enforced by sanctions', such as the rule that 'employees are not to smoke within 100 yards of any acetone vat' (Lewis 1986, 102). To this list Lewis adds three other kinds of rules that are not conventions: first, rules that are 'threats or warnings issued by some authority or power to control the behaviour of a class of people against their own preferences' (such as a rule in a POW camp against gathering in groups of more than six, with a violation to be punished by ten days on bread and water) – here 'one's incentive to obey is the same whether or not the rest obey' and for that reason cannot be a convention; second, rules that 'codify regularities' by prescribing 'behaviour for each agent which may go against his own preferences…but which answers to the preferences of everyone else concerned' (e.g. stating what ought to be done to keep promises or reciprocate benefits) – here again there is an absence of a co-incidence of wants; and third, rules that 'are enforced with sanctions so strong that one would have a decisive reason to obey even if others did not' (Lewis 1986, 103-4).

An example used by Lewis to discuss the last of these non-conventional rules is instructive. He refers to the use of standard notation in pages of some academic journal. Here, something that may start out as a convention – for there is a need to adopt a common notation, and, by precedent, this becomes the favoured possibility for reaching an equilibrium – can over time become a rule, enforced by strong sanctions (in this case, rejection from the journal) and, in doing so, cease to become a convention (Lewis 1986, 104). This is instructive because it helps us to see that (at least on a rational reconstruction of these terms) whereas conventions depend on regularly exercised behaviours, rules do not – although, as noted above, by no means does the fact that conventions depend on regularly exercised behaviours mean, for

these theorists, that they are equivalent to mere regularities; all it means is that there is a need to account for what this 'something more' in the case of convention amounts to; rules, on this view, are then the 'something else', whereas conventions are 'something more', than mere regularities.

Of course, distinguishing rules from conventions does not mean that they cannot both be in play in the context of some activity. Take, for instance, games. According to Lewis, games are certainly governed by rules, which he calls 'listed rules' (Lewis 1986, 104). Violation of these rules would be 'decisive evidence of inability or unwillingness to play' the game (Lewis 1986, 104). Interestingly, Lewis says these rules are 'conventional', but he also adds that 'they are not the only conventions in the game. Any group of players will develop understandings – tacit, local, temporary, informal conventions – to settle questions left open by the listed rules' (Lewis 1986, 104). Here, Lewis is alive to the possible different uses in terminology, for he says that 'We may call these understandings rules – unwritten rules, informal rules – if we like; but we would also be inclined to emphasise their differences from the listed rules by saying that they are not rules, but only conventions' (Lewis 1986, 105). This is important for Lewis to mention because when we hear the often-mentioned dictum that 'language is governed by rules', we might – mistakenly, according to Lewis – come to think of rules 'that have been codified by some authority' (analogous, say, to the listed rules of chess, codified by the International Chess Federation), and we may thus come to look for rules 'enforced by sanctions, formal or informal', or 'of rules that are mentioned in teaching or criticising the use of language' (Lewis 1986, 107). If we have this view of rules when pronouncing the above dictum, we are bound to be disappointed: we will not see rules being referred to in the teaching of language, and we will not see rules mentioned in criticisms of certain uses (we may, of course, but this will be a peculiar or specialised sort of teaching or criticising). If, however, we allow for an understanding of the rules governing language to be one that sees those rules as conventions, then we will avoid that doomed search. We will then, says Lewis, find 'regularities in behaviour', but not just regularities, but rather 'expectations and preferences regarding verbal behaviour, and in expectations regarding these expectations and preferences' (Lewis 1986, 107). It is worth keeping in mind Lewis's comments here, and his efforts at distinguishing rules from conventions (and thus finding a distinction between games and language), for we will be returning to the claim that language, like games, is governed by rules in

Part II of this thesis (especially in Chapter 8; in doing so, we will also be critiquing this particular understanding of games: see Chapter 7).

C. Replies to Lewis

There have been, over the years since his book, many important replies to Lewis – and, simultaneously, alternative accounts of the nature of conventions. Chapter 16 of this thesis will examine one of these replies – a recent one by Andrei Marmor (2009) – in some detail.

One of the most famous and influential replies was offered by Margaret Gilbert.[21] Gilbert finds two basic problems with Lewis's account: first, she argues that it is possible for conventions to exist even in the absence of any underlying regularly exercised behaviour, i.e. in the absence of anyone conforming to the convention anymore – her example here is that of there still being in existence a convention of sending of thank you notes after a dinner party in a certain group, although no one in the group sends thank you notes anymore; and, secondly, she argues that Lewis neglects the normative dimension of conventions.[22]

In terms of the second point, Gilbert says that 'conventions in Lewis' sense do not seem apt to give rise to the "ought" judgments typically associated with conventions as ordinarily conceived' (Gilbert 1989, 354). Gilbert defines conventions in the following way: it is common knowledge in population P that all members of P have intentionally and openly manifested their willingness jointly to accept a certain principle of action.[23] As articulated recently by Michael Rescorla (2011), Gilbert's view is that 'our everyday concept of a social convention is that of a jointly accepted principle of action, a group fiat with respect to how one is to act in certain situations' (Rescorla 2011, 28).

Gilbert's account of the nature (including the normativity) of conventions is inseparable from her social metaphysics. She has a view, developed in detail since *On Social Facts*, that understands social activities – including seemingly simple activities, such as walking together (see Gilbert 1990) – as constituted by certain norms (i.e. norms that dictate what it means to walk together in this culture, e.g.

---

[21] The initial reply was in *On Social Facts* (1989), chapter 6. See, more recently, Gilbert 2008.

[22] Gilbert also argued that Lewis's account was 'singularist' (in Gilbert 1989) and 'individualist' (in Gilbert 2008), but the two criticisms above are most pertinent for present purposes. Gilbert also doubts that coordination games offer a promising starting point for the analysis of conventions.

[23] The formulation here comes from Latsis 2005, 716.

walking together in cultures with gender equality may mean walking side-by-side, whereas in cultures without gender equality, the man might be required to walk ahead of his wife). Armed with such metaphysics, she then proceeds to offer a theory of how persons come to acquire knowledge of, and commit to, such constitutive norms.[24]

It is unusual to think of conventions as attitudes detached from underlying regularly exercised behaviours. Even, however, where there is no underlying regularly exercised behaviour identified, as in Gilbert, the attitude in question is said to be regularly exercised and, in Gilbert's case, is cashed out in terms of commitment to constitutive norms – in that sense, the attitude retains the picture of the passivity of the person (regularly committing to the norms that 'own' the normative, so to speak). In other words, however it is articulated – and the crux of the conventions debate really concerns how (deliberatively and normatively) demanding the attitude in question ought to be understood[25] – the attitude is typically said to be regularly exercised. In that sense, although something 'more' is added to the merely regular (or something 'else', as in Gilbert's case), the appeal is in the end still made to regularising that something 'more' or something 'else' (i.e. the appeal is made to regularising the normative). In that move – tht is, in the regularisation of the attitude – we witness the identification of the normative with the ordered and patterned. In the literature of conventions, because conventions are often said to be dependent on an underlying regularity, persons are positioned passively in two forms: first, by reproducing of the 'same' behaviour (this being the content of the underlying regularity, i.e. the repeated behaviour); and second, by regularly exercising the attitude (often latched on top of the underlying regularity). As we shall see in more detail later in the thesis, this passive positioning of persons is deleterious to our understanding and description of the abilities persons need to participate in the social.

---

[24] For a more detailed discussion of Gilbert's notion of joint commitment, including an attempt to apply it to competition law, see Del Mar 2011b.

[25] Seamus Miller 2001 refers to a 'standing intention to some shared collective end', which may remind us of Gilbert's notion of joint commitment (for a discussion, see Rescorla 2011, 29), and in that sense can be said to be quite demanding. An example of something less demanding is Ruth Millikan's 2005 theory, arguing that 'a convention is a pattern of behaviour reproduced within a population due largely to weight of precedent' (Rescorla 2011, 29). Millikan pits her account precisely against Lewis who, from her perspective, 'tried to establish how social order emerges from the rational decisions of individual agents' (Rescorla 2011, 30).

Chapter 2. Norms

The most common determinants of appropriateness and inappropriateness referred to by the governance view are not conventions, but rather norms and rules. Indeed, we have already seen how Lewis, in discussing the normative dimension of conventions, turned to articulating how conventions were (in practice, likely to be) like norms and sometimes even like rules. The above discussion of conventions, then, is offered more in the spirit of completeness, with more emphasis being placed here on norms and rules. Conventions were important to consider because they are often identified as the closest to so-called mere regularities, or as most dependent on them (*pace* Gilbert). Of course, as we shall see, norms (unlike rules) are also often thought of as reliant on there being an underlying regularly exercised behaviour, with something 'more' added (where that something 'more' is itself a regularly exercised attitude), but it is generally unusual to think that the something more added by conventions and norms is the same, i.e. it is often said that whereas conventions typically require the co-incidence of wants and interests, and thus are likely to emerge precisely because of the mutuality and mutual interdependence of interests, norms (in this respect like rules) are more likely to either emerge or be imposed as constraints on wants and interests. This is only rough, however, for, as we shall see, there is enormous variety in the uses of the concept of a norm. The distinction alluded to here between conventions and norms is returned to in Chapter 4.

The etymological root of the word 'norm' is *norma*, meaning a carpenter's square. According to Nicholas Dent, the term 'norm' is itself of 'fairly recent coinage', unlike the term 'normal', which has been more familiar for longer (Dent 2005, 662). Of course, of importance here is the difference between the norm and the normal, so it is noteworthy that the terms share a common root. Dent himself defines a norm as that 'which signifies either the average or usual level of attainment or performance for an individual or group; or, and more usually in philosophical discussion, a standard, rule, principle used to judge or direct human conduct as something to be complied with' (Dent 2005, 662), thereby suggesting that the term itself can either mean norm in the strict sense or norm in the sense of normal. Of note, also, is that in defining norm, Dent refers to a range of other terms – standard, rule, and principle – that many philosophers (as we shall see in more detail when we discuss rules) would want to distinguish between.

As general definitions go, however, Dent is not alone in so defining 'norm'. Stephen Finlay, in a recent article reviewing recent literature on normativity also defines norm as 'a criterion, standard or rule to which something either does or does not conform and against which it can be compared or measured' (Finlay 2010, 332). This is similar, but it is worth observing that Dent in fact includes more in his definition than Finlay, because for Dent being able to serve as a criterion is not enough – what is also needed is that the norm 'direct human conduct as something to be complied with.' In Dent's definition, then, there is appeal to a difference between something that just happens to serve as a standard against which things are measured, and something *designed* to serve as an instrument for both evaluating and directing behaviour.

A. Von Wright's Taxonomy of Norms

A popular port of call for philosophers setting out to investigate norms – and articulating this something 'more' that allegedly distinguishes the norm from the normal – is the work of G.H. Von Wright. Von Wright has a very expansive view of norms (it is the umbrella term for him, not only for all things normative, but for all patterned things), but the virtue of his approach is that he pays particular attention to this difference between the norm and the normal. That Von Wright's use of the term norm is so broad can be seen from his assertion that we can divide norms into three kinds of laws: the laws of the state, the laws of nature and the laws of logic (Von Wright 1963, 2). Laws of nature are 'descriptive': 'they describe the regularities which man thinks he has discovered in the course of nature. They are true or false. Nature does not, except metaphorically, "obey" its laws. If a discrepancy is found to exist between the description and the actual course of nature it is the description, and not the course of nature, that must be corrected' (Von Wright 1963, 2). Already here, Von Wright is offering a number of markers of difference between laws of nature and the other two kinds of laws: the direction of correctness (or what contemporary philosophers tend to call, following John Searle, 'the direction of fit');[26] the applicability of a judgement as to truth or falsity of the law; the sense in which laws may come to exist (in this case, they are 'discovered', suggesting the laws are there, whether we know it (yet) or not); and the different possible meanings of 'obligation'.

---

[26] For a useful overview, including pertinent quotations from Searle, see Humberstone 1992.

Von Wright asks us to contrast these characteristics with the two other kinds of laws: first, the laws of a state; and second, the laws of logic. The former are prescriptive as opposed to descriptive. They are 'laid down' rather than 'discovered' and they are laid down as 'regulation for the conduct and intercourse of men' (not as descriptions of nature; again, we have reference here to a designer). Unlike laws of nature, they have no truth-value; rather, 'their aim is to influence behaviour' (Von Wright 1963, 2; as with Dent, we see reference here to the 'aim' of the designer). When violations occur, the usual reaction is not to change the law, but rather to 'correct the behaviour of men', though Von Wright recognises that sometimes 'the authority alters the laws – perhaps in order to make them conform more to the capacities and demands of "human nature"' (Von Wright 1963, 2). The contrast with laws of nature could not be, according to Von Wright, more stark or straightforward.

Less straightforward, says Von Wright, is the third category, that of the laws of logic. Certainly, laws of logic 'prescribe' – in this case, they 'prescribe how one ought to think and calculate in order to think and calculate correctly' – and, in so doing, they 'provide a standard whereby to judge whether people think correctly or not' (Von Wright 1963, 4). Looked at more closely, however, Von Wright is not entirely satisfied with seeing the laws of logic as prescriptive. He wants to say something stronger. He recognises that laws of logic are also, in some sense, descriptive, but not descriptive in the sense that they tell us how people think, but rather that they tell us 'how the logical entities are constituted' (Von Wright 1963, 4) – it is only by reference to those laws that we can know what thinking or calculating is. Perhaps, then, he suggests, we ought to think of the laws of logic neither as truly descriptive or prescriptive, but rather as 'determinative' – as determining what it is to think and calculate (Von Wright 1963, 6).[27]

Von Wright's classification of the above three kinds of laws (all under the umbrella of norms) does not stop there. Rather, he introduces several others kinds of norms. For example, Von Wright calls a rule the 'main type of norm' (Von Wright 1963, 6). In speaking of rules Von Wright introduces an image that we will encounter often in this thesis, i.e. the image of games. Here is his clearest statement in this respect:

---

[27] It needs to be stressed that this language of determinacy is Von Wright's own language; the constitutive strategy that Von Wright appeals to here is addressed later in the thesis.

Playing a game is a human activity. It is performed according to standardised patterns, which can be called moves in the game. The rules of the game *determine*, as I shall say, these moves or patterns – and thereby also the game 'itself' and the activity of playing it. We could say that, when viewed from the point of view of the game itself, the rules *determine* which are the correct moves, and when viewed from the point of view of the activity of playing, the rules *determine* which are the permitted moves. It is understood that the moves which are not correct are prohibited to players of the game, and that a move which is the only correct move in a certain situation in the game is obligatory when one is playing the game. (Von Wright 1963, 66; emphasis added)

For present purposes, the most noteworthy aspect of the above is the sense in which Von Wright refers to rules as determinants of what is correct and what is incorrect.[28] Noteworthy, as well, is that Von Wright goes on from the above passage to apply the image of games to that of language: language, he says, is rule-governed in the way that games are rule-governed, though with the difference that with language we do not yet know all the rules or cannot describe them with absolute completeness and precision (Von Wright 1963, 67). We will be coming back later (in Chapter 8) to views about how language is rule-governed – typically, by implicit rules that we have not yet made explicit (and thus somewhat like, if not completely like, a game; we have seen the same point being made by Lewis above). It is also noteworthy, though it also shall not be dwelt on at this point (but see Chapter 11 on computation), that Von Wright goes on to speak of thinking itself as playing with symbols (Von Wright 1963, 67) – what we see here is that a certain understanding of rules (of what they are capable of, e.g. determining what is appropriate or not, determining what it is to think or not) trickles down to a view about cognition, i.e. that we cognise – or even play, in Von Wright's case – in accordance with rules.

Apart from rules, Von Wright also speaks of customs as a kind of norm, though he identifies both norm-like and non-norm-like aspects of customs. Customs are norm-like 'in the sense that they influence conduct; they exert a normative pressure on the individual members of the community whose customs they are. The existence of this pressure is reflected in the various punitive measures whereby the community reacts to those of its members who do not conform to its customs' (Von Wright 1963, 69). But customs are also, he says, 'species of habits…a regularity in an

---

[28] 'Permission' is understood as the same thing, but from a different perspective, i.e. the perspective of a participant in the game – so participants do not necessarily think about what is the correct thing to do, but about what is permissible or obligatory, but both permissibility and correctness are determined by the rules.

individual's or community's behaviour', something 'that is regularly done' (Von Wright 1963, 68). The habitual aspect of customs, then, does not mean that they cannot serve as standards for criticism of behaviour, i.e. it can be the case that when the custom is not done (either exceptionally or habitually), it is 'regarded with disapproval' (Von Wright 1963, 68; this adds the something 'more' needed to make it arguably distinguishable from a mere regularity). Von Wright's example in this respect is that 'it is a custom of my country, but not of Anglo-Saxon countries, to thank the hosts or the heads of a family when the meal is finished' (Von Wright 1963, 68). Here, the something 'more' that customs have that mere regularities (or perhaps mere habits) do not is that 1) they influence conduct; and 2) that this pressure on conduct is reflected in criticism from deviations. Von Wright does not elaborate as to how we can tell which customs have this something 'more' and which do not, but his definition does imply that some do and some do not, i.e. that it is stable, and established, in any community, which departures (from patterned ways of doing) will be regarded with disapproval and which not.

For completeness, it should be added that towards the end of his classification of kinds of norms, Von Wright introduces another tripartite division of 'types of norms'. Here, the division is into rules, prescriptions and directives (Von Wright 1963, 15). Rules we have already met, and we have already observed that Von Wright thinks of rules along the model of rules of a game, which he extends to rules of grammar (see also Von Wright 1963, 15), and it is also rules in this sense that he thinks may be closest to laws or rules of logic and mathematics (Von Wright 1963, 15). Prescriptions are 'commands, permissions and prohibitions, which are given to or issue to agents concerning their conduct', and these include the laws of the state (Von Wright 1963, 15). Finally, directives are what Von Wright calls 'technical norms': 'they presuppose the ends of human action' and they constitute the 'necessarily relationships of acts to these ends' – in other words, they are the necessary means to the ends they presuppose (Von Wright 1963, 15). All these norms have contents, by which Von Wright means 'that which ought to or may or must not be or be done' (Von Wright 1963, 71).

The 'ought' that Von Wright has in mind here is one that he wants to say is addressed to 'doing' rather than 'being'. For Von Wright, norms – whether we think of them as classified under laws of the state and laws of logic, or as classified as including customs but divided primarily into rules, prescriptions and directives – all

address conduct. Nevertheless, it is important to see that however they are addressed (i.e. to agents), norms are situated in an environment that is presupposed to have determined what it is that agents can (choose to) do – norms here function as conditions of possibility for what can be done (whether this be playing a game or thinking), with agents addressed, and positioned, only in a passive sense, i.e. of being confronted with certain available choices.[29]

For Von Wright, then, it is true that norms address agents and thus doing rather than being, but that is in a context in which it assumed that there are certain activities or ends to be pursued, the point being that the conditions for participating in those activities or ends is determined by the norms. The tendency here is to merge the epistemology with the ontology: there is the ontology established by norms (where the norms are conditions of possibility of participating in certain activities or pursuing certain ends) and there is the correlated normative epistemology, i.e. coming to know and thus being guided by and regulating one's own and others' conduct by reference to those norms.[30] The effect of such a move is that it positions persons as incredibly passive, static and rigid: their choices are identified for them, and what is relevant (and thus possible while at the same time appropriate) is always and already established and never created by persons. The effect of this positioning is that we miss certain important abilities that persons need to participate in social life, for, on this picture, it is assumed that the only abilities that persons need, or at least the most important ones, are the ones that will enable them to acquire knowledge of, and appropriately internalise, those norms. This same point can be put another way: given that it is assumed or posited that the norms establish what is there to be explained, it is, thereafter, only those abilities that internalise those norms that are picked out as important for persons' participation in social life. As a result of this, once again, we miss certain other abilities that are important for us to participate in social life.

---

[29] There are in fact two different senses of 'choice' here: there is the choice of playing a game, and the choice of logic, the point being that whereas we can choose to be regulated by the rules of a certain game we cannot, on Von Wright's analysis, choose to be regulated by the laws of logic; this is precisely the difference between metaphysical and practical necessity.

[30] A further example of a statement along these lines can be gleaned from Ota Weinberger, a philosopher sympathetic to Von Wright's analysis of norms, who says: 'The ontology of norms is based on the thesis that norms determine action… action is not only part of the content of norms, but norms themselves should be understood as factors which determine human action and the constitution of human societies' (Weinberger 1985, 313; again, the language of 'determinacy' is Weinberger's own).

Von Wright's discussion of norms has been highly influential in analytical philosophy, but it is by no means the only account. It will be useful to very briefly consider some other accounts of norms, not only in the philosophical literature, but also in the social sciences and in psychology.

B. Norms in the Philosophy of Social Science

Reference to norms is common in the social sciences, and in philosophical reflection on social science. The philosophical literature itself includes a variety of uses. Todd Jones, for instance, in a series of articles on norm-based explanations of human behaviour, has suggested that norms tend to be used in one of two ways: first, as one of a group of 'cognate terms like custom, convention, tradition, culture' that generally point to 'what's done', where what turns 'what's done' into a norm is simply how widespread it is (Jones 2009, 3); and second, as a term used 'to refer to a certain set of social causes of behaviour', as when we say 'that Mohammad will not hold hands with his betrothed in public because of the norms of his society forbid it' (Jones 2009, 3). Jones adopts the second way, i.e. the 'social cause' use of norm, which in earlier work (in a paper in 2006), he defines as indicating that 'a person or persons did Y because they thought lots of others do Y and / or expect Y to be done' (Jones 2006, 257). Neither of these uses refers to the otherwise commonly encountered idea in philosophical discussions that norms, in addition to being regularities, are standards for the evaluation of behaviour and / or (more usually 'and') guides intended to direct behaviour, but it instructive that even when defined without these additions, Jones requires them to be mediated by the 'thought' of what others do and / or expect to be done (what 'thought' amounts to is not further defined).

A more familiar use (given what has been analysed above) of the term is one that defines a norm, as Jon Elster does, as 'an injunction to act or to abstain from acting', where some of the injunctions are unconditional ('do X', 'do not do Y'), and some conditional ('If you do X, then do Z', or, 'If others do X, then do X'; Elster 2007, 354). Typically, on this reading, norms are distinguished from conventions on the basis that the latter can be enforced through the sheer self-interest of the agent, without any action by others, whereas the former operate through the emotions of shame in the norm violator and of contempt in the observer of the violation (Elster 2007, 355-7). Thus, for example, a convention can emerge that on the second day of a conference everyone will gravitate to the seats they sat on yesterday: it is in

everyone's self interest to find a seat, and also to find it quickly, so it is easier to do what everyone else is doing, i.e. following what was done yesterday. That convention can be contrasted with the norm to always wear black clothes at a funeral, where it need not be in everyone's interest to do so (for example, it might suit someone to wear sports gear, if say that person plans to go exercising afterwards). Further, not wearing black may certainly arouse contempt in others at the funeral, and shame in oneself for not having worn black (the point being that these emotions would be absent from violations of the convention, e.g. me sitting in a different seat to yesterday, unlike everyone else).

David Dequech (2009) offers a neat summary of the above way of reading norms vis-à-vis conventions, when he says that norms 'are social standards of behaviour and / or thought that a) indicate what people should or should not do or think under some circumstances; b) are at least in part enforced upon individuals by external sanctions and c) are internalised by some or even many individuals' (Dequech 2009, 72, with the proviso that I may so internalise a norm that my following it no longer requires the threat or promise of external sanctions), while conventions have 'at least two additional characteristics', the principal two of which are: 1) 'when followed consciously, a convention is followed at least in part because other people are following it, and not, or not only, because there is an external pressure to comply; and 2) it is to some degree arbitrary, in the sense that a hypothetical alternative could replace the actual pattern' (Dequech 2009, 73).[31]

A brief, though instructive for present purposes, philosophical discussion of norms as explanations of behaviour in the social sciences is provided by David Henderson. Henderson's general definition of norms is that they are 'dispositions to co-ordinated patterns of action and evaluation within some group of people' (Henderson 2002, 327). Spelling this out some more, he says that 'to characterise a people as having such-and-such a norm is to characterise a pattern of action exhibited in the "fitting circumstances", and to say that members of the group have dispositions to conform to such a pattern and to evaluate action (or actors) with respect to its conformity' (Henderson 2002, 327-8). The latter part of this spelt-out definition is important, for Henderson wants to stress that although a norm is a pattern, it is also

---

[31] The arbitrariness condition can be questioned, for it may also initially have been arbitrary whether black is chosen as the colour to wear at funerals – the key difference seems to be more a matter of the kind of attitude expressed, and the kind of emotions felt, when someone is evaluated to have departed from the norm or convention.

more than a pattern: what is added is that behaviour is monitored for conformity, and is sanctioned if it does not conform. In calling behaviour norm-governed one is saying something not only about how persons are likely to act, but also about how they are likely to react (positively when conformed to, and negatively when not; Henderson 2002, 328). For example, 'to say that there is a norm of avoiding eye contact between strangers in a certain large city is to say that folk there typically do not make eye contact with people they do not recognise and with whom they do not have business, and that, *ceteris paribus*, they act as though such contact is generally bad – perhaps impolite or intrusive, and perhaps a symptom of mental illness, rural hick-icity, or various other moral or psychological failings' (Henderson 2002, 328). What we see here is not only dependence on an underlying regular behaviour (a pattern), but also that that the something 'more' that is added (for there needs be a certain kind of reaction built in) is itself a pattern (in Henderson's terms, a disposition). In this respect, Henderson's view is not too far from a view like Jones's (2006 / 2009) definition of norm as social cause, though Henderson does not require any mediation by a 'thought' as Jones does, but rather refers to a disposition to 'react', meaning that the likelihood of reacting is enough to capture the something 'more'. Jones, it should be said, is not necessarily trying to capture or assert there is something 'more', but he builds in what others who do wish to say there is something 'more' would also build in, i.e. precisely that mediating 'thought' (again matters are made difficult by the fact that Jones does not quite say what he means by 'thought').

As indicated above, Henderson does not stop with the general definition, but goes on to contrast his account of norms to a number of other accounts. Henderson's concept of a norm acknowledges that it 'need not make for the objective correctness or appropriateness of the actions that conform' (Henderson 2002, 328). This, he says, is in stark contrast to some other accounts, e.g. accounts of the norms for rationality where (according to Henderson) the focus is on '*the* norms of rationality', such that a gap opens up between those norms on the one hand, and, on the other hand, norms that a people might have regarding how to reason, i.e. as 'tendencies to reason in certain ways and to evaluate reasoning accordingly' (Henderson 2002, 328). On the view Henderson wants to distance himself from, the norms of rationality can be discovered as 'independent of how we, or any people, are presently (or at any give time) disposed to reason and evaluate reasoning' (Henderson 2002, 328). For Henderson, this is to claim too much about norms. Henderson wants to avoid

postulating some timeless, universal, set of norms that are said to determine the objective appropriateness or correctness of, say, what it is to be rational.

This is an important difference with the constitutivist, but it is also important to see that there are aspects of the constitutivist's way of thinking that Henderson does not quite escape. This we can see in how Henderson criticises the move towards something constitutive or objectively correct. He says, for instance, that the norms he is concerned with 'do not themselves have what it takes to constitute what is correct' and then he adds 'the actions or thoughts to which they give rise can be uniformly wrong' (Henderson 2002, 329). What is puzzling here is from what perspective those actions or thoughts can be 'uniformly wrong.' Certainly, Henderson's view is one that does not make the same claims as to the status of the norms that he thinks may govern some community. Henderson's view builds in the possibility of change in a way that the constitutivist does not. But the problem is that Henderson's view is still one where it is imagined that there is an established (even if defeasible) order of what is appropriate and inappropriate – the addition of it being subject to change is, though an improvement on the constitutivist view, not quite enough in escaping normative determinacy (i.e. in there being something according to which those actions or thoughts would be 'uniformly wrong').

The second difference that Henderson identifies between his and other views of norms – e.g. Brandom's – is that, according to Henderson, many accounts of norms (especially as 'rules') have 'to do with computational algorithms. Talk of rules commonly is of a piece with the idea that one who follows a rule, runs a program. This requires that the one who follows a rule has a set of rule-like representations instructing the agent (or some subagential component) what to do in what conditions' (Henderson 2002, 300). What Henderson does not accept is what this view suggests, namely that 'to describe a norm one should be able to specify a program that the relevant agents are running' (Henderson 2002, 300). For Henderson's norms to work – as 'co-ordinated dispositions to action or thought, and to evaluation' – 'agents must have learned something', but the difference is that 'there need be no suggestion that this learning was a matter of acquiring or internalising a tractably computable algorithm' (Henderson 2002, 300). Henderson wants to avoid a computational picture, but what he wants to replace the computational picture with is one that still focuses on describing 'the capacities that humans have for complying with social norms' (Henderson 2002, 300). Henderson, then, still subjects the description of how persons

learn and act to a pre-supposed order of norms. In other words, the norms are imagined first, and then agents are described as learning and acting in accordance with norms (though without being mediated by some computing of what the norm requires). This positions persons passively and rigidly vis-à-vis that which could emerge as normatively relevant, with a pernicious effect on how active human cognition is described (for more details of the issues here, see Part III of this thesis).

C. Norms in the Sociological and Psychological Literature

Before we move on to rules, let us briefly consider how norms feature in the sociological and psychological literature (of course, the aim here is not comprehensiveness, but just a glimpse at uses). Many of the uses in the sociological literature refer to norms as expectations, though they differ in terms of other features. George C. Homans, for instance, argues that 'a norm is a statement specifying how a person is, or persons of a particular sort are, expected to behave in given circumstances – expected, in the first instance, by the person that utters the norm. What I expect of you is what you ought to do' (quoted in Opp 2001, 102), though this is clarified to mean expectation in the form of a demand about how you ought to behave, not an expectation about how you will behave (Opp 2001, 102). To this basic sense of 'normative expectation' one might add a requirement for a sanction, or one might add a requirement for a sanction plus behavioural regularity. Thus, one can end up with, as Karl-Dieter Opp does, the following conditions of the existence of a norm: 'a norm exists if there is a behavioural regularity, if deviation from this regularity is sanctioned and if there is a normative expectation that the behaviour is to be performed' (Opp 2001, 103).

A similar definition to Opp's is that offered by Jack Gibbs (1965). Gibbs says that 'a norm involves 1) a collective evaluation of behaviour in terms of what it ought to be; 2) a collective expectation as to what behaviour will be; and/or 3) particular reactions to behaviour, including attempts to apply sanctions or otherwise induce a particular kind of conduct' (Gibbs 1965, 589), though what is missing here (in comparison with Opp) is the requirement for the norm to be supported by an underlying regularly exercised behaviour (i.e. it is not explicitly stated that the object of the regular collective expectation is itself regular), and what is included (that is not in Opp) is a predictive explanation on top of a normative one (it is said that the

expectation in question is both an expectation in terms of what it ought to be and in terms of what it will be).

One of the other ways in which norms are used in the sociological literature is as means to ends, though this itself can be understood in a variety of ways. Thus, Leonard Broom and Philip Selznick, for instance, argue that 'norms are blueprints for behaviour, setting limits within which individuals may seek alternate ways to achieve their goals' (quoted in Gibbs 1965, 586; see Broom and Selznick 1963). David L. Miller, on the other hand, says that 'the purpose of a norm is to inform us how it is possible to apply technology in achieving a certain end without interfering with the achievement of other desirable ends' (Miller 1951, 138). Although it may not be obvious at first, there is a distinction in the two definitions, for in Miller's case there is a reference to norms 'informing us', whereas in Broome and Selznick norms are more like conditions of possibility (i.e. they are more like the rails that one does not see but within which one acts). In Broome and Selznick, norms are not subject to choice, but are instead the conditions of possibility for choice: it is within a certain space created by norms that persons choose and act (on the basis of their choice). In Miller, norms are instruments to be used for certain ends, and thus are subject to choice, though choice here is still imagined as a matter of choosing which norms out of a catalogue of available norms – persons are seen as having more choice, but still presented with a certain range of possible means they can adopt towards a certain range of possible ends.

Finally, a norm in the sociological literature may be defined as a form of explanation, a kind of generalisation made by a theorist, though ideally one that reveals something about how persons behave. Thus, for instance, Nicholas and Carol Pollis define a norm as a 'standardised generalisation that defines an expected and/or appropriate range of behaviours given a particular class of relevant social situations' (Pollis and Pollis 1970, 231). A norm here is a tool of sociological explanation that identifies 'explicit or implicit factor[s] influencing individual experience and behaviour' (Pollis and Pollis 1970, 231). The focus here seems to be on the attitude or mental process (expectation) thanks to which a regularly performed behaviour is reproduced, though there is some ambiguity in the definition, for Pollis and Pollis also seem to allow for an explanation based solely on identifying what is 'the appropriate range of behaviour' in certain social situations (the 'or' in the first of the above-mentioned quotes suggests an independence of this range from expectations).

In summary, and it is of course a very rough one, sociological references to norms tend to refer to expectations attached to underlying regularly exercised behaviours. These expectations are typically said to be collective expectations of what ought to be done (though sometimes also, at once, predicting what will be done), which may be supported by sanctions, i.e. the person not only expects and predicts what another will do, but is also likely to punish them if they do not that. This is the primary use in the literature, but the literature also contains reference to norms as ways in which a community, society or culture (or any other kind of grouping) determines the field of choice and / or constitutes the very content of choice. Finally, norms can also be referred to as forms of sociological explanation of 'an appropriate range of behaviour' in certain social situations, where the appropriateness in question may be more or less dependent on expectations (the object of which is precisely that behaviour) being regularly exercised. In all these cases, although there is sometimes effort to distinguish the normal from the normative, the normative is nevertheless depicted as a regularly performed order, typically in the sense of an underlying regular behaviour that is the object of an attitude (e.g. an expectation, with or without likelihood of sanctioning) that is itself regularly exercised.

In terms of the psychological literature on norms, according to Leslie Smith, the usage tends to be one that refers to 'normal or average performance' (Smith 2006, 10). Where they are not so limited, norms refer to directives from other people in authority – as was the case in Stanley Milgram's famous experiments (see Smith 2006, 12; Milgram's experiments are briefly discussed in Chapter 18). Smith herself, in a collection of papers entitled *Norms in Development* (Smith and Vonèche 2006), bemoans the dominance of the 'non-normative' meaning of norms, and based on what she refers to as the 'normative understanding of norms' (following partly Brandom and partly Von Wright; see Smith 2006, 19-21) she argues for a need for a normative meaning of norm to be used in psychology, especially developmental psychology (other papers in the collection are likewise so inclined).

In addition, experiments in social psychology typically make a distinction between two senses of the term 'norm': first, so-called descriptive norms, i.e. that which is commonly done; and second, injunctive norms, i.e. that which is commonly approved and disapproved, where the distinction is said to lie in these two norms having separate sources of motivation (see Kallgren et al 2000, 1002; see also Cialdini et al 1990). Here, the hypothesis (which has been tested) is that making

salient what is commonly approved and disapproved in some community (e.g. with experiments where a confederate disapproves of someone littering) is more likely to result in a person avoiding doing that thing than simply making salient what is commonly done. We will be returning to these experiments when we discuss Christina Bichierri's work (in Chapter 15).

This completes the brief survey of the use of the concept of norms, principally in the philosophical literature (but mentioning also uses in the sociological and psychological domains). We have seen that where theorists are alive to, and want to establish, a distinction between the normal (or the merely regular, ordered, and patterned) and the normative, what is appealed to most often is some kind of attitude, whether this be a mediating attitude (some 'thought' or some 'expectation of what ought to be done') or a reactive attitude (the disposition to react critically, as in Henderson's case), and sometimes a complex attitude that contains both (is both mediating and at least potentially reactive).

The pertinent observation from the perspective of this thesis is that these attitudes (either the mediating or the critical attitude, or a complex attitude) are themselves typically said to be regularly exercised (e.g. in the form of abilities, dispositions or tendencies). In fact, the overlap with regularity can come in three forms: first, that the actual behaviour in question is one that is regular (i.e. there is an underlying regular behaviour that is reproduced); second, that the mediating attitude (e.g. the expectation) that has as its object that behaviour is regularly exercised; and third, that the reactive / critical attitude that also has that behaviour as its object is also regularly exercised (whenever a departure from the expected behaviour is perceived). In many theories of norms all three are present, i.e. there is a regularly performed behaviour to which is attached the regular exercise of mediating and critical attitudes.

These overlaps between the normal (or the regular) and the normative ought to put us on alert. They should put us on alert because we ought not focus exclusively or even dominate our understanding of the social, or of its normative dimension, by reference to that which is regular. Just as important as the regular is that which is irregular, and, thus, just as important as the abilities that enable us to latch on to the regular are those abilities that connect us with the irregular. These latter abilities are missed by the exclusive or even dominant focus on the regular. In that sense, there is a limitation on the role that norms should be play in constituting the object of

explanation (of the social, or the social in its normative guise) and in the explanation of that object.

Chapter 3. Rules

Let us proceed a little more inductively than in the above two Chapters, and begin with some examples of rules:

- A pawn reaching the eighth rank must be exchanged for a piece (Black 1962, 95);
- No smoking in classrooms (Black 1962, 106);[32] and
- White moves first (Ganz 1971, 7).

Contrast these with the following, some of which may look like rules, but are not (they used by Max Black and Joan Safran Ganz as contrast cases):

- The dealer at bridge always bids first (Black 1962, 106);
- Black moves first (when 'Black' is taken to be the surname of a particular player; Ganz 1971, 13);
- The law of gravity (Ganz 1971, 72);
- Announcing an attack on the queen by saying 'queen' (Black 1962, 131);
- Turn left at the traffic light and walk straight for three blocks (Ganz 1971, 84);[33]
- Close the door (Ganz 1971, 87, given as an example of a command or an order); and
- The ball is 'live' as soon as the whistle blows (Ganz, 100).

Given these examples and contrast cases, what have been the attempts by theorists to find a way of formulating the difference? Let us consider some of these attempts, beginning very briefly with Black (1962), but thereafter focusing on Ganz (1971).

A. Black on Rules

According to Black, rules 'must be about the kind of thing that human beings can do' (Black 1962, 107). Presumably, this would immediately exclude the above example

---

[32] Though, according to Black, when this is pronounced in a certain way, it may be an order: Black 1962, 107.

[33] This is given as an example of directions or instructions, but Ganz 1971 notes, at 85, that it may be a rule – it depends how this statement is used.

of a law of gravity: humans comply with this law whether they want to or not, so it does not make sense to ask whether it is something they can do. Further, the formulation of the rule must 'indicate with respect to those actions whether they are required, forbidden or permitted' (Black 1962, 108). Clearly this excludes the second of the above contrast cases, for 'Black moves first' (where Black is the surname of the player) does not require, forbid or permit anyone to do anything. It will also probably exclude the fourth of the above counter cases, as when we say that that some group of players announce 'queen' when they attack the queen (analogously to announcing check when attacking the king). Of course, in that case, we would need to know more about the group in question, but phrased as it is, it suggests a custom or observance rather than a rule. We get a little bit of help from Black as to why when he says that 'the word "rule"…alludes to a spectrum or continuum of cases, in which pressure of one sort or another is exerted through the formalised means of the pronouncing of rule-formulations' (Black 1962, 123). Thus, in the case of announcing 'queen', we would presumably need to know whether a formalisation existed, and whether this formalisation played a role in exerting pressure to conform; of course, 'pressure' itself is a vague term, but here it needs to be read in light of Black's understanding of the components that make up a rule formulation, namely: '1) a class of human 'actions'; 2) a class of persons who perform those actions; 3) an indication of whether those actions are demanded, forbidden, or permitted' (Black 1962, 137). In other words, for Black, if there was a formulation that asserted that the action of saying 'queen' when attacking the queen was, for persons who were playing chess, demanded or permitted, and this formulation was one that exerted pressure on the players, then we could say in the above case that we were dealing with a rule.

This is not yet entirely satisfactory (it does not, for example, rule out some of the other contrast cases), so let us turn to Ganz (1971), who offers the most comprehensive and sustained analysis of the nature of rules in the philosophical literature (and an account that, unfortunately, is often neglected).

B. Ganz on Rules

The merit of Ganz's approach is that he slowly works through the various examples and contrast cases (accumulating general features of a definition of rules as he goes along). Take the difference between saying that 'White moves first' and 'Black moves first', where the 'Black' in the second formulation refers to the surname of the player.

Syntactic criteria alone, as well as generality or universality, are not, Ganz argues, sufficient to distinguish the cases, but he does think that 'Black moves first' is a statement that can be shown to be true or false, whereas, he says, there is no sense in which 'White moves first' can be said to have truth-value (Ganz 1971, 24). Furthermore, 'Black moves first' is not an inscription that is followable (Ganz 1971, 27) and it is not an inscription that specifies 'what counts as a correct (or incorrect) procedure for an activity' (Ganz 1971, 50).[34] Furthermore, Ganz asserts at one point (and repeats numerous times thereafter) that not only must rules specify what counts as correct (and incorrect), but they must also specify *the one* correct procedure, and thus not be one among various strategies that could be adopted (Ganz 1971, 52). So far, then, we have the following elements of rules: they are 'followable linguistic entities having no truth value which specify what counts as correct procedures for activities' (Ganz 1971, 51), where the correct procedures that are specified are the only correct ones and no others are allowed.

It was mentioned above that Ganz's use of prescriptive includes the specification of a (probably *the one*) correct procedure for an activity. Added to this element is the idea that the rule serves as the basis for the favourable or unfavourable evaluation of behaviour. Part of what it means for a rule to be prescriptive, then, is that when one conforms to it, this is a basis for a favourable evaluation of that behaviour, and when one does not conform to it, this is a basis for an unfavourable evaluation (though it should be noted that Ganz asserts that this is a necessary but not sufficient condition, because other phenomena may serve as such a basis, and yet not count as a rule; see Ganz 1971, 53).

Still more is said by Ganz concerning the prescriptiveness of rules. Thus, rules 'are prescriptive not only in their being the basis of evaluation of behaviour by specifying what counts as one correct procedure for activities, but also by being tools used by an authority for calling upon those, who have the obligation to fulfil rules made by the authority, to behave in certain ways' (Ganz 1971, 59). Ganz does make room here for two different sources of 'prescriptive force': 1) where the prescriptive force 'derives from the relation of rules to activities; one must fulfil rules if one is to merit favourable evaluation of behaviour with respect to that activity' (Ganz 1971, 65); and 2) where the 'authority behind the rules' yields the prescriptive force of the

---

[34] Ganz explicitly asserts that his use of 'prescriptive' as opposed to 'descriptive' just is that prescriptive rules specify what counts as correct and incorrect (Ganz 1971, 51).

rules, at least as long as 'the authority has both the right to specify correct procedures and the right to require one to conform to those procedures' (Ganz 1971, 65).

Despite making room for this alternative source of prescriptive force (i.e. one not dependent on an authority behind the rules), it has to be observed that the presence of such authorities features prominently in Ganz's analysis. Even where he does not refer to authorities explicitly, he refers to an authoritative procedure for adopting the rules. Thus, he argues that some method appropriate to the situation must have been employed which bestows the status of a rule on the relevant linguistic entity (Ganz 1971, 65-6). Rules, to be rules, must be appropriately 'adopted', i.e. 'adopting a rule…is for the appropriate means to have been taken by the appropriate people so that an utterance or inscription specifies what counts as correct or incorrect procedure for some activity' (Ganz 1971, 66). This is a necessary condition for the existence of a rule, according to Ganz (1971, 66).

Putting this requirement for adoption together with the ones above, we have a definition of rules as 'appropriately adopted prescriptive linguistic entities having no truth value' (Ganz 1971, 72), where we need to supplement the notion of prescriptive with the above-mentioned features (i.e. the basis for an evaluation of behaviour, either favourable or unfavourable, where the prescriptive force comes from either the way the rules constitute the activity or because of the authority behind the rules).

Ganz's analysis of the difference between rules (as understood above) and the laws of science is instructive. First, as already indicated, Ganz notes that the laws of science are not followable (Ganz 1971, 72). Second, although it is possible for some linguistic entity to be both a law of science and a rule (Ganz 1971, 77), they would cease to be either a law of science or a rule for different reasons, i.e. a law of science is disconfirmed by a case that does not conform to it (all other things being equal), whereas a rule remains a rule despite the absence (even wholescale) of conformity to it.

It is important to see that Ganz pushes the definition of rules as far away as he can from regularity: what they require does not need to be performed (at all – not just regularly), and their existence does not depend on either a (regularly exercised) mediating or a (regularly exercised) reactive attitude (though it can enable either attitude to be exercised at any one time: one can use the rule to guide oneself *ex ante* or one can use the rule to evaluate behaviour *post facto*). Perhaps as a result of pushing rules this far, his accounts ends up characterising rules as linguistic entities

and emphasises the role of an authority behind the rules (or at least an authoritative procedure). What needs to be seen here is that this greatly restricts what can count as a rule, and thus also restricts the role that rules can play in the establishment of an object of explanation and in featuring in the explanation itself.

We have not yet distinguished rules from a number of other phenomena. Consider, first, customs and habits.[35] Rules differ from these in that there is less claim to correctness associated with customs and habits: 'if a man fails to perform habitual actions he is more likely to be surprised or be thrown off stride than to judge himself as having acted wrongly' (Ganz 1971, 80). Of course, if that person did evaluate himself negatively on the absence of having performed what was a habitual or customary action, then we might say, according to Ganz, that 'he has adopted a rule for himself' (Ganz 1971, 81). The general point is that the 'criterion for rule behaviour is the attendant assessment made, not its regularity' (Ganz 1981, 81). Here, Ganz does make a reactive attitude necessary (in order to distinguish a rule from a habit) – but what he emphasises about it is not that it be regularly exercised, but that in the moment that it is exercised it results in the adoption of a rule. What is vital to see here is that Ganz does not say that the exercise of a reactive attitude reveals that the rule was always being followed; rather, in reacting the way we do, we make a rule for ourselves. This is vital to see for it avoids the implication of passivity vis-à-vis the normative, this being an implication made in the appeal to implicit rules (i.e. rules precisely 'revealed' by the exercise of a critical attitude).

Consider, next, directions – as in the fifth contrast case above. These, says Ganz, are 'recommendations for how to proceed to your destination'; they need not be followed 'in order to get where one wants to be' – they are at best 'suggestions for how one could get to the destination' (Ganz 1971, 84). What is missing from directions is the claim to correctness, especially in the sense of there being only one possible procedure. When there is a rule, and not a direction, then the 'means for getting there are specified with the expectation that the performer will go by the route stated. Only such performance will count as satisfactory' (Ganz 1971, 84). Again, what is important to see here is that although this does help in clarifying the distinction between a direction and a rule, it also has the effect of narrowing what can

---

[35] In fact, Ganz refers to not just customs and habits in this context, but also: set forms, observances, uniform or established courses, customary ways, practices, procedures, paradigm cases, wonts, prototype procedures, settled forms, and conventional practices (Ganz 1971, 80).

be referred to as a rule. The normativity of a rule is not the looser normativity we might associate with aims or ends (recall Railton's characterisation of that in the category of the evaluative), but the narrower normativity associated with the legislating of one correct procedure.

What about commands (recall the second last counter case, i.e. 'close the door')? These are different to rules, says Ganz, because 'rules are still rules after having once been followed, but commands are not commands after having once been obeyed' (Ganz 1971, 88; directions and instructions may also be short-lived in this way). There is a sense, then, in which rules are more long-standing and not confined to immediate use, i.e. they remain in force (they remain the relevant basis of assessment) until un-adopted (Ganz 1971, 89).[36] Consider, also, the following differences between orders and rules: 'rules are made by appropriately adopting them', whereas 'orders are given by uttering them' (Ganz 1971, 93);[37] 'orders are directed at people', whereas 'rules are directed at activities… rules specify procedures for activities' (Ganz 1971, 93); and 'rules are conditional; they act as critiques for behaviour only if certain antecedent conditions obtain' (Ganz 1971, 94). In an arresting image, Ganz – following Black in this respect – likens orders to shots, as opposed to rules, which are like fences. As Black put it, 'an order is like a shot fired at a poacher, a regulation like a fence to keep anybody off the grounds' (Black 1962, 119).

Ganz also distinguishes rules from principles. Unlike rules, 'principles don't specify the procedures for action'; instead, they 'provide the ideology, justification, motivation and the like for procedures' (Ganz 1971, 96). Of course, rules can also be used to justify action, but unlike principles, their 'content is not ideological but methodological… principles are a matter of policy; rules are a matter of required procedures…' (Ganz 1971, 96). Rules 'determine' (as Ganz himself puts it: 1971, 98) the correct procedures for an activity.

Recall that in determining the correct procedure, the rules also exclude any other alternative means. According to Ganz, this also distinguishes rules from 'models', which Ganz calls 'exemplary rather than prescriptive' (Ganz 1971, 99).

---

[36] Their 'remaining in force', however, is not the same as being the object of a regularly exercised attitude. They can remain in force 'on the books', as it were.

[37] The difference lies in the manner of issuing orders and rules – notice that this is a better criterion that saying that rules, unlike orders, are impersonal or indirect, for orders could also be impersonal or indirect (Ganz 1971, 90).

Models, he adds, 'are rather like directions and instructions' – they 'show a scheme or procedure which could be followed but need not be followed' (Ganz 1971, 99). Methods that are not rules provide suggestions, or offer strategic advice, or give clues about efficient means, but they are not prescriptive in the way that rules are (Ganz 1971, 99). In this respect, again, 'maxims' are not rules, for they are merely advisory and miss the sense of 'compulsoriness' of rules; 'while rules specify what counts as the one correct way to perform an activity, maxims specify either the most strategic, most practical, most accepted, or most well thought of and the likely, way to perform an activity' (Ganz 1971, 102).

According to Ganz, then, 'Utterances and inscriptions called rules are appropriately adopted, prescriptive, conditional linguistic entities having no truth-value' (Ganz 1971, 104). We have, above, looked at the meanings of these elements, including the sense of 'prescriptive', under the umbrella of which Ganz places the prescriptive force and compulsoriness of rules. For present purposes it is important to see that Ganz requires rules to determine the one correct procedure for doing something: it is this that then functions as the basis for evaluating behaviour – either favourably, if complied with, or unfavourably, if not – and it is this that provides the sense of 'requirement' or 'compulsoriness' that distinguishes rules from a whole variety of other phenomena that purport to give guidance or advice (such as instructions, directions, maxims, models, etc, which all leave room for possible alternatives). It is also important to see that in saying that rules determine the one correct procedure for doing something, Ganz finds himself emphasising the role of authorities legislating the rules or, at the very least, the following of appropriate procedures for laying down the rules. Finally, it is important to see that rules, on Ganz's definition, are independent from regularly exercised behaviours: they are, as it were, rules 'on the books', linguistic entities issued authoritatively by authoritative institutions. They are neither revealed by critical attitudes (though they can enable them – this does not mean that critical attitudes cannot be exercised without them, just that they can be exercised with them) nor need they be mediated by any regularly exercised attitudes – the point being that the rules are not dependent on being regularly taken as guides (and thus not describable as followable by default).

But now a curious thing comes to surface once rules are defined in the way Ganz does. Having made rules independent of regularity in the way Ganz does, it must be realised how narrow the scope of the account is, narrow at least from the

perspective of the explanatory ambitions of the governance view. No reference to 'implicit rules' is permitted, for rules are linguistic entities adopted appropriately (most probably by authorities recognised as authoritative). For similar reasons no exercise of a critical attitude is proof of any underlying rule. Further, there is no scope and no need for trying to turn rules into regularly exercised attitudes, e.g. by presenting them as forming the content of abilities or dispositions. Instead, what we have in referring to rules is a kind of resource (in the form of one rule or a set of rules) that persons may use, contingently, in order to guide themselves or evaluate themselves or others. There may be circumstances in which persons wish to, or feel a need to, control what is to be done in some very specific way; they can then treat as unchangeable a certain prescribed means for attaining a certain end, adopting an attitude that there is no other means permitted. But there is no claim here that such treating is either *a priori* necessary or necessary in some practical way (i.e. it constitutes a defeasible order composed of an underlying regularly exercised behaviour and attached regularly exercised attitudes).

Ganz offers a very plausible account of rules, but his account also points to the narrowness of the concept: we should see it not as determining appropriateness and inappropriateness in some sense that establishes an *a priori* necessity or a default order, but as a contingently useful resource used in circumstances where, for one reason or another, persons may wish to treat certain ways of doing things as unchangeable. To discuss these matters, however, is already to begin to delve into criticism of the use that the governance view makes of rules, and this goes beyond the scope of this Chapter (more about these issues is said in the next Chapter).[38]

As we shall see in later Chapters, especially in Parts II and III, the concept of a rule is used more broadly by the governance view than Ganz does.[39] The concept of a

---

[38] It needs to be observed that by no means is this a thesis about rules (in Ganz's sense) and the various contingent uses that can be made of them. If it was a thesis about rules, it would need to delve into much greater detail concerning various contexts, e.g. adjudication, where rules are relied on and used in specific kinds of ways. Certainly, the thesis wishes to note the usefulness of rules (again, in Ganz's sense), but it also argues that the normative (in general, but especially the normative dimension of social life) ought not to be exhaustively understood on the basis of how rules are used.

[39] It should be added here that sometimes 'rules' are used an umbrella term for the patterned, e.g. Allan Grimshaw asserts: 'All students of human social behaviour…see that behaviour as persistent, patterned, non-random, rule-governed…' and later in the same paper, 'linguists have always directed attention to regularities…and consequently, to rule-discovery procedures' (Grimshaw 1980, 790). This is too general a definition for present purposes (for it does not allow for a distinction between rules, norms and conventions, let alone between rules and regularities), but it is useful to mention it, for it does indicate how widely uses of the term 'rule' are. It can also be added here that there is no consensus in the philosophical literature on which term (rule or norm) is more fundamental. Often, one

rule is often appealed to as distinct from a mere regularity, but nevertheless not adopted – not created by an authority (including a person for themselves). When this is done – as when rules are said to govern some domain without us realising it, with us only being able (and even then not always) to reveal underlying rules in exercising a critical / reactive attitude – the advantages of Ganz's picture are lost sight of. In other words, precisely by confining rules to linguistic entities adopted, typically issued by a body recognised as authoritative, Ganz manages to rescue the activity of cognition (which, as we shall see, will be important in helping us see some of the other abilities that we need to participate in social life). Put differently, in confining rules in the way he does, Ganz shows us how we can see the normative as independent of the regular, but only if we position ourselves actively vis-à-vis the normative. Our appeal to a rule – including our creation (or 'adoption' in Ganz's terms) of a rule – is an active one: it is a contingent treating of something as required and unchangeable (as when a certain procedure is treated as the only correct one). Any attempt to insert regularity back into a rule – e.g. by appeal to implicit rules followed tacitly (and we will see many examples of this in Part III) – re-introduces passivity, and turns what is contingency (i.e. the contingent treatment of something as unchangeable) into a kind of necessity, either of an *a priori* kind (as with rules of logic) or a practical kind (as is said often of games). What needs to be avoided, then, if we are to rescue activity, is the refusal to think of rules as constituting some underlying or designed order and one that we slavishly follow. Adopting Ganz's definition is part of how we can achieve this, but more needs to be said about how we ought to approach normative resources. Saying more, and thus also contrasting how conventions, norms and rules are understood on the governance view on the one hand, and from the second-person perspective on the other hand, is the task of the next Chapter.

---

is used to explain the other (and vice versa): e.g. Akeel Bilgrami defines a rule as a norm that guides or constrains behaviour or thought (Bilgrami 2005, 824). Again, an attempt is being made here to rationally reconstruct the uses of these terms, and thus not treat one as explicable in terms of the other.

Chapter 5. A Rational Reconstruction

We have seen in the above Chapters that the terms 'conventions', 'norms', and 'rules' have been used in a wide variety of ways. We shall see this, too, in the literature to be discussed in future Chapters. In the first Section of this Chapter an attempt is made to rationally reconstruct these terms vis-à-vis each other. The point of doing so is to make clearer where the thesis stands in relation to these very commonly referred to phenomena. It will, in addition, be important to show that as much as the thesis distances itself from both the creation of an object of inquiry via conventions, norms and rules, as well as from any explanation dominated by reference to conventions, norms and rules, it recognises that they are important for our understanding of the social, including its normative dimension.

The thesis distances itself from assuming or positing any of these as necessary, and also necessarily shared, and it does so because the effect of this – given the connection between conventions and norms and regularity, and the tendency to regularise rules – is to squeeze out any space for that which is irregular, which, from the second-person perspective, we can see is so important for social life and our ability to participate in it. Nevertheless, the thesis acknowledges that we can sometimes treat that which a convention, norm or rule requires to be done as unchangeable – in that sense, the convention, norm or rule can itself seem necessary, including necessarily shared. As shall be argued below, however, and also in later Chapters, it is important to see that this is a form of active and contingent treating of something as unchangeable, i.e. treating as unchangeable for certain purposes or in certain contexts that which the convention, norm or rule requires to be done. The purposes in question are typically justificatory or explanatory; or, put differently, the contexts in which a convention, norm or rule can seem necessary is in a context of justification or systematisation. The contexts of justification and systematisation, however, are not the only contexts in which we can have a normative experience, an experience concerning something that ought to be done. That experience is also possible in what is called in this thesis the context of appreciation. The second Section of this Chapter elaborates on these points, which are also referred to in later Chapters. The point, in short, is to make room for the phenomenology of normative necessity, but not to let it either create the object to be explained or exhaust the kind of explanation that can be given (of social life, including its normative dimension).

A. The Connection with Regularity

In the literature discussed above, conventions and norms often overlap with regularity in three ways: first, the behaviour they require is regularly performed; second, the mediating attitude (the attitude that enables guidance) is regularly exercised; and third, the critical / reactive attitude (which enables both self- and other-regulation or evaluation, sanctioning departures from the convention or norm) is regularly exercised. Put differently, persons regularly perform what conventions and norms require, and they regularly use what those conventions and norms require as guides to their conduct and as regulation of their conduct and the conduct of others. Norms Conventions and norms, then, form a kind of underlying defeasible regular order: persons will continue to perform the required actions, be guided by those requirements and refer to them in evaluating their own and others' behaviour, until and unless given reason to do otherwise. Thus, when it comes to conventions and norms, persons are describable as stocked with defeasible mediating and critical / reactive attitudes that enable them to act in accordance with the norms and conventions that are then, typically, presented as conditions of possibility for participation in social life, and at once granters of meaningful participation. Indeed, it is often said that only behaviour that is expected normatively is intelligible, and being intelligible is also appropriate; the bounds of possibility, intelligibility and appropriateness are often said to hang together by the tapestry of required behaviours identified and determined by norms and conventions.

As has been noted above, it is this connection of conventions and norms with regularity that also marks their limitation from the second-person perspective, i.e. conventions and norms, insofar as they are used to create the object of explanation (social order) and insofar as they then feature as explanation of how that social order is maintained (via the internalisation of that which ought to be done according to conventions and norms), both squeeze out the possibility of irregularity featuring in the object of explanation and in the explanation itself (or, more accurately, in the description of the abilities we need to participate in social life, including its normative dimension). Put a little differently, if we focus exclusively on conventions and norms in our pursuit of the social, including its normative dimension, we exclude from that which needs to be explained, and from the explanation, all that which is irregular, unpatterned and disordered, which, as the second-person perspective shows, is so

important for social life, including its normative dimension. Further, this focus on conventions and norms – and thus the focus of the governance view in general – is one that does not recognise the kind of normative experience we have when social life is seen from the second-person perspective: as we shall see in later Chapters, the governance view simply cannot see, let alone recognise the importance of, abilities to appreciate and attend to others in interaction understood from a second-person perspective.

Although conventions and norms share a good deal in common, they are also subtly different. Unlike conventions, norms are not addressed to the self-interests of agents. Instead, norms are addressed to agents in order to try to assist them in overcoming their self-interests and act in a way that fosters what is commonly imagined to be in the common good (on this way of putting it, there are certain values that are constitutive of that community, and these values have yielded norms, which one can conceive of as means that further those ends). Norms are needed because otherwise agents might not feel compelled to contribute to the common good, and instead exploit others for their own ends. Conventions, on the other hand, have their own kind of compelling force, for they are addressed to agents already disposed to finding a solution to a certain kind of problem: one where some action is required, no matter what it might be, e.g. persons need to sit down somewhere in a lecture theatre, or persons need to drive on one (rather than randomly either) sides of the road (it does not really matter where they sit – but notice that if it did, e.g. someone was hard of hearing or seeing and needed to sit up front, there might be a norm in place to forestall able persons from sitting up front so as to allow those in need to sit there).

Another way to mark this distinction between conventions and norm is to point, as does Margaret Urban Coyne (now Walker) in her paper entitled 'Mapping the Normative' (1981), to a 'normative continuum'.[40] Coyne's continuum stretches from the 'normative-critical' to the 'normative-sympathetic', where the former refers to what we impose because we want to control and restrain the self-interest of

---

[40] From the perspective of this thesis, it is also interesting to note that Coyne begins her paper by marking a distinction between 'the patterned and the unpatterned' (Coyne 1981, 331). The former, she asserts, 'includes behaviours which have a typical form across like situations or occasions; the latter, behaviours which are either responsive to unique situations, or at least not consistently typical across like situations' (Coyne 1981, 331). For our purposes, it will be important to see how that which is here thought of as unpatterned, and not worthy of being the object of explanation (at least when it comes to the normative), is, on the contrary, of vital importance (though to say this is already to characterise in ways we should avoid, i.e. part of the point is to avoid this division between the ordered (and the good, the predictable, the meaningful) and the chaotic (the bad, the random, the meaningless).

persons, and the second what tends to emerge more naturally from those interests (Coyne 1981, 336). Coyne contrasts conventions and rules (for our purposes these are better thought of as 'norms') and notes that 'in conventional behaviour the sources are multiple and complex in a way that the motive and incentive for rule-conformity need not be' (Coyne 1981, 335). In the case of rules (or, for us, 'norms'), these are often advantageous, or work particularly well, where persons cannot be trusted to want to conform, i.e. where the conforming behaviour in question may be sometimes seen as at odds with the self-interest of persons. Conventions arise where the very existence of a pattern (of conforming behaviour) makes it the case that it is in the interest of the person to conform. As Coyne says, 'without general conformity to some typical format an enormous number of mundane pursuits and activities would become intolerably complicated, exhausting, tedious, or disruptive' (Coyne 1981, 335). Sometimes, sanctioning measures may be used, but they 'are very often unnecessary, for people can be counted upon to look out for their own interests, to know certain facts about behaviour in their group, and to make certain assumptions about other agents with whom they must contend (Coyne 1981, 335). In short, once a convention is established, there is often 'no good reason in the absence of unusual circumstances not to repeat the pattern oneself, and to do so with the confident expectation that others will do likewise' (Coyne 1981, 335). Conventions, then, can be distinguished from norms for conventions appeal to the interests of persons whereas norms operate precisely to defeat those interests.

Yet another, though in the end closely related, way to discuss the difference between conventions and norms is to point to different attitudes that their represent. Conventions might be said to represent a comparatively weak normative attitude, i.e. an expectation that certain behaviour (in such-and-such a context) will be repeated. If such expected behaviour is not performed, the person with the expectation may be disappointed, but they may not go so far as to expressed criticism, though they may ask, or at least look, for an explanation. One could illustrate this with the example of a group of friends who establish a 'convention' to meet in a local pub on a Sunday at a certain time. When one of them does not show up, as expected, they may be disappointed, and they may look for an explanation, but they may not go so far as to

criticise that person.[41] Certainly, it is possible to have an expectation that has no normative tinge – would only be accompanied by surprise if something that is expected does not occur – but this seems different from an expectation to which a reactive attitude of disappointment is attached.

This weak normative attitude – including a weak reactive attitude – arguably characteristic of conventions can be compared to the stronger, though not yet as strong as could be, attitude that characterises norms. Persons do not expect (or, better, not only expect) something to be done according to a norm – they require it, and when one someone (in their judgement) deviates from the norm, there typically follows not merely disappointment, but rather criticism. Put another way, a person who is the victim, or even just the observer, of something they adjudge to be a deviation from a norm is not going to merely ask for (or look for) an explanation: they will demand a justification. They will then exercise judgement as to whether that justification merits recognition as a justification. Thus, for example, if someone rushes up to sit in a seat reserved for those who are hard of hearing or sight, but does not seem to suffer from those disablements himself, then a justification may be sought, and if it is sought and one is given – e.g. that they need to stretch their legs in the front row – then this justification will be evaluated as either sufficient or not. This, clearly, is a normative attitude – including a reactive attitude – that is stronger than that which arguably characterises conventions.

If this line of reasoning is to be followed, then rules might be said to relate to an even stronger normative attitude. It could be said, for instance, that the attitude is not merely one of expectation or requirement, but a demand. Where, furthermore, the demand is not met, the relevant reaction will not be either disappointment or criticism, but the imposition of a sanction. What will be looked for is not an explanation or a justification but, instead, a defence. Thus, for instance, the rule that prohibits extracting evidence from torture will be a rule that, if adjudged to be broken, will

---

[41] In their recent helpful paper, where this example appears, Nicholas Southwood and Lina Eriksson (2011) argue that there is nothing normative about conventions. This, however, imposes a rather strict interpretation of the normative as requiring the exercise of criticism when there is deviation. Arguably, requiring or even just looking for an explanation might be classified as a normative attitude, even if a very weak one. Southwood and Eriksson also argue that conventions are unlike norms for they are behaviour dependent and desire dependent. Again, here, their argument is helpful, but could perhaps be softened by an acknowledgement of degrees of dependence on behaviour and / or desire, i.e. norms need not be completely independent of behaviour and / or desire; just less dependent in those ways than conventions are. The language of degrees suits the universe of normative attitudes better than any zero-sum game.

result in the imposition of a sanction, unless a defence (if a defence at all is to be allowed) will be shown (e.g. imminent danger to human life as a whole).

Of course, we need not use the language of conventions, norms and rules to point to such differences in attitude, but certainly, if we are to stick to those terms, then a good way of rationally reconstructing their meaning might be to point to such differences in attitude.

This thesis does not place a great deal of emphasis on these differences in attitude. What it does place emphasis on is the difference between conventions and norms on the one hand, and rules on the other hand, in relation to the connection with regularity. In Ganz's words, as analysed in Chapter 3, rules are 'appropriately adopted, prescriptive, conditional linguistic entities having no truth-value.' What they require need not be regularly performed, and the rules need not be regularly used as guides, nor regularly appealed to in criticisms of behaviour. In that sense, rules are disconnected from regularity. They are not, however, disconnected from order, and that is also why they are one of the explanatory resources of the governance view.

There are, in fact, two ways of seeing rules as related to order. The first way is to assert that (at least some of) these rules constitute *a priori* orders, constituting the very possibility of certain domains (e.g. the rules of logic as constitutive of thinking or rationality), or they constitute designed orders (such as games, designed by authorities), such that persons generally comply with what those rules require (even when they do not know that they do, as is often said to be the case for *a priori* orders) or they learn the rules (especially those designed, and thus e.g. the rules of games) and, after a time, come to internalise them and follow them (precisely in the sense of regularly exercising a mediating and critical reactive attitude – thereby becoming a kind of defeasible regular default order as well as being *a priori* or designed). This is the way that the governance view tends to view rules. In other words, it uses rules to 1) set up the object of explanation (either because the rules constitute the very existence of that domain or because they do so in a practically necessary way); and 2) having set up the object to be explained in that way, they then regularise the rules, i.e. they describe persons' abilities to participate in those domains (whether *a priori* or designed, i.e. whether metaphysically or practically necessary) as based on the internalisation (sometimes even without awareness or knowledge) of those rules (e.g. as we shall see later, with respect to grammar, the rules are said to 'grow' in the minds of persons until, at least, they 'master' the language).

This way of seeing rules related to order is characteristic of the governance view, and it is this way of seeing rules should feature in our understanding of the social that this thesis argues against. The second way, and the way that is endorsed from the second-person perspective, is to view rules (still conceived as linguistic entities, along the lines of Ganz) as contingently useful resources. The idea, in brief, is to disassociate rules with any kind of necessity, whether *a priori* or designed. At the same time, however, it is recognised that that which rules (are understood to) demand can sometimes be treated as unchangeable – in that sense, rules can seem (we can experience them as) necessary. In other words, the way that rules are related to order from the second-person perspective is as a contingently-built order on top of something that is not itself necessarily ordered, patterned or regular. Again, this thesis does not argue against rules as such, or even against the orders that rules may compose (as when we see a set of rules related to one another in a consistent way). Rather, the point is not to restrict, from the beginning, our understanding of the social, including its normative dimension, as being necessarily ordered, such that, thereafter, the only abilities we see as important for our participation in social life are ones that tend to be explained on the basis of the internalisation of the rules that are said to constitute those necessary orders. Order cannot be seen to be the underlying base or the explanatory target: it is just something contingently built on top of something that, in itself, is neither necessarily ordered or necessarily disordered (but rather a mix of the regular and the irregular).

On the governance view, then, there is a tendency to first secure the ontological status of rules. This is done either by appealing to their alleged *a priori* status (as is often said to be the case for the rules of logic) or by claiming that the rules were designed by authorities (where the authorities are legitimate, i.e. recognised as having the power to issue rules). The positing of *a priori* status to rules is at odds with the understanding of the social, including its normative dimension, from the second-person perspective. The positing of rules designed by authorities is not at odds with that perspective, but it becomes divorceable from it when, with the ontology of the rules secure, it is asked how persons come to 1) act in accordance, regularly, with the rules; 2) know the rules, so that they can not only act in accordance with them, but also regularly follow them (use them regularly as guides and as standards for criticising behaviour); and 3) know the rules by revealing them incrementally (this being especially the case with *a priori* rules, the idea being that

over time, as we dig deeper and refine our instruments, we can come to reveal – make explicit – the underlying implicit rules that, say, govern our thinking or rationality). When these moves are made, rules come, in the end, to be assimilated into a kind of regular order, regular because persons are said to regularly comply with the rules, and, over time, come to regularly guide themselves and regularly criticise themselves and others on the basis of those rules. There are, then, two kinds of order here: the *a priori* or designed (ontological) order, and then the (ideal) epistemological order where persons reveal, come to know, and come to internalise (use as guides and standards for criticism) the constituents of that order.[42]

We shall see many instances of this way of treating rules (sometimes, in the literature we will be discussing, the same idea is referred to under the term 'norms') in Parts II, III and IV of this thesis, though each of these focuses on different aspects of the governance view, i.e. Part II focuses on arguments that seek to establish the (metaphysical and practical) necessity of rules; Part III on the effect of such necessity on how minds are described as learning and working (for, on the back of such a view, minds are described precisely as learning these rules – or learning abilities that happen to result in conformity to the rules – such as to be able to regularly perform what is required, while also regularly exercising the relevant attitudes); and Part IV on the sense in which these rules are necessarily shared (in any group, community, society, culture, etc, that is allegedly necessarily governed by them).

One way, then, from the perspective of this thesis, of understanding the problem with this way of appealing to rules, i.e. beginning with ontologically secure rules (*a priori* or designed), and then describing persons as learning to act in accordance with them, is that it makes the epistemology secondary to the ontology. In other words, securing an order of rules that determine what is appropriate and inappropriate, and then proceeding from there to use that order as the epistemological and behavioural target – i.e. as the ideal target not only for behaviour (ideal behaviour

---

[42] In addition to 'designed orders' one could speak of 'emergent orders', as where two persons, in the course of interacting, create and negotiate a kind of order between them (this being different to designed orders for these depend on the prior positing of rules by authorities). The notion of emergence is not at odds with the thesis, but the crucial thing is whether the theorist who appeals to emergence thinks that an order emerges, such that this order then becomes the regular default one, governing the interaction of the parties from then on (until either or both the parties are given reason to do otherwise) or whether the theorist allows for the possibility that what emerges may be 1) indeterminate, or 2) if an order, then an order in the sense of a pool of contingently useful resources to be appealed to, if need be, by either party (and not something that acquires default applicability). If the theorist allows for the second possibility, then the second-person perspective becomes possible. If not, despite the welcome appeal to emergence, the theory comes to take on the garb of the governance view.

on this view is one that complies with the established order) but also for one's knowledge and one's attitudes (the content of which becomes the requirements of those rules, both as guides and as standards for criticism) – results in persons being depicted as perpetually passive vis-à-vis that which can emerge as normatively relevant, and as static and rigid compliers with and followers of pre-existing / pre-established orders.

Whether, then, appeal is made to conventions or norms, or to rules, on the governance view, the normative is a kind of order: it may be an order that emerges from interactions between persons, going on to establish itself as the default order that is composed of underlying regular behaviours and accompanying regularly exercised mediating and critical reactive attitudes, or it may be an order that is *a priori* or designed, composed of rules, which is then ideally regularly complied with, persons coming to 'reveal' or 'discover' hidden or implicit rules, and going on to regularly use them (in the forms of abilities, dispositions, tendencies, sensitivities, etc) as guides and standards of criticism. And it is precisely because it is imagined to be an order – where, ideally, the ontology and epistemology map onto one another – that persons come, as a result, to be depicted as passive vis-à-vis that which can emerge as normatively relevant, and as static and rigid participants, their passions restricted to ever effectively policing each other to act in accordance with the requirements of the order.

B. Contingent Ordering

As noted above, this thesis does not argue against the utility of the concept of order when it comes to understanding the social, including in its normative dimension. It does, however, sound a note of caution. Let us not, it pleads, begin with and thus construct an explanatory target of order – whether composed of conventions and norms, or of rules, or some mixture of these. Let us, instead, acknowledge the limitations of the appeal to the concept of order and note that order can be contingently built on top what is itself neither ordered or disordered, but a mixture of the two. Further, let us acknowledge that we can have the experience of normative necessity, but that this experience tends to be confined to certain contexts – as we shall see below, to contexts of systematisation and justification – and that there can be other kinds of normative experiences, and thus also other kinds of abilities relevant to participating in social life in its normative guise, that do not require the mediation of

that which order is said to be composed of (again, whether conventions, norms or rules, or some mix of these).

Strictly speaking, from the second-person point of view, we ought to resist the opposition between (intelligible, appropriate, meaningful) order and (chaotic, random, meaningless) disorder. It is better not to begin with an ontologically secure (either *a priori* or designed) order, which thereafter becomes the epistemological target. Instead, it is advisable to flip the priorities around: to establish, at bottom, not a default order to be departed from only with good reason, but rather indeterminate relations (neither ordered or disordered but, at best, a mix or balance of the two) on top of which orders may be built in contingent fashion (for certain purposes, when the need arises). Put slightly different, from the second-person perspective the indeterminate is the underlying foundation, with the determinate built contingently on top. Thus, there is no opposition between the determinate and indeterminate; instead, one remains the *status quo* (the indeterminate) until a need arises for the determinate to be built contingently, and for limited purposes, on top.

'Indeterminacy' can be an ugly word, referred to when one cannot think of what else to call that which one finds difficult to articulate, and so this thesis will use it only sparingly. The crucial point is that what we see as important for social life, including its normative dimension, from the second-person perspective is interaction that is not itself always and already ordered, regular and patterned, but instead, one that is, instead, a mix or balance of the regular and the irregular etc. In that sense, opposing determinacy and indeterminacy may be preferable to opposing order and disorder, especially when we understanding indeterminacy to include a mix or balance of the regular and irregular, with the determinate (the strictly ordered) contingently built on top for certain purposes. As we shall see in more detail later, interaction, from the second-person point of view has an irregular rhythm. In it, persons relate to each other in a tentative and responsive way, though sometimes punctuated by, or informed by, certain normative resources (whether conventions, norms or rules) being used in an active and contingent way. Put differently, and if we accept the rational reconstruction attempted above, in interaction from the second-person point of view we may sometimes exercise the kind of attitudes, and have the kind of normative experiences, that conventions, norms or rules can be associated with. We can, in addition to that, for certain purposes, treat the behaviour that is the object of those conventions, norms or rules as unchangeable, and thus experience

those conventions, norms or rules as necessary. But none of that is foundational or fundamental. Instead, at bottom is an interaction that is tentative and responsive, and as important as these convention /norm / rule-based attitudes / skills / experiences are, they are but one aspect of that interaction.

On the governance view, what we are here calling contingently useful normative resources are seen, instead, as determinants of appropriateness and inappropriateness that are necessary (ontologically and epistemologically) and necessarily shared. From the second-person perspective, the social, including its normative dimension, ought not to be cashed out exclusively or even dominantly by reference to determinants of appropriateness and inappropriateness, i.e. by reference to orders constituted by conventions, norms and / or rules. Nevertheless, the second-person perspective does not ignore conventions, and norms and rules. Instead, it characterises them differently, i.e. not as determinants of appropriateness and inappropriateness, but as contingently useful resources. Similarly, the second-person perspective does not ignore the benefits of *ordering* (with the emphasis on the *–ing*): it just does not presuppose, and treat as ontologically secure order; instead, it highlights the sense in which persons can, at various times for various purposes, appeal to, including create and then maintain, orders, or even experience those orders as necessary, and thus treat the relevant required behaviours as unchangeable.

From the second-person perspective, persons can, for various purposes, treat normative resources (whether conventions, norms or rules) as applicable to something they and / or others do, but that this 'treating' is an active and creative exercise, and not something passively and rigidly exercised by regular default (performed automatically or mechanically, unless given reason to do otherwise). When persons relate with others, and to some extent also the environment, they can do so – and typically do so – without automatically or mechanically framing, or imposing some order of salience, on the other or the environment. Persons, then, are not accurately describable as necessarily and sufficiently pre-stocked with certain ways of framing or making salient, such that they go on to apply such-and-such rules relevant to such-and-such a frame or domain of salience, departing from this only when given reason to. They are not, as it were, carriers of the automatic or mechanical application of determinants of appropriateness and inappropriateness. Other persons or the environment do not confront persons precisely in the way those persons are pre-prepared or pre-structured to see them. Persons are not stocked with abilities,

dispositions, tendencies, sensitivities, commitments or standing intentions that they immediately exercise unless given reason to, as if they automatically and mechanically saw other persons and the environment in certain determinate ways (again, unless and until they are surprised or jolted into seeing them in some other way). Instead, persons relate with others and the environment, in the first and foundational instance – this being precisely their default relation – in a way where other persons and the environment appear indeterminate to them. Other persons and the environment are approached by persons, in the first instance, in a way that can create what might be relevant to see; what may be relevant is not, in the first instance, given: it is suspended, with a person going on to either create what might be relevant or indeed, if need be, reverting to a familiar resource to guide them.

The details of interaction from this second-person perspective cannot be delved into at this point, but it is important that something more is said about the 'active treating' referred to above, i.e. the active treating of some behaviour as required (in a way that may appeal to certain familiar normative resources). There are two contexts in which this active treating typically occurs: contexts of 'justification' and 'systematisation'.

A context of justification is one where we are interested in making a claim or assertion about how something is or ought to be, and thus we scan the environment for grounds upon which we might be able to make those claims, e.g. we look for certain properties (and treat them as being properties) that enable (on the basis of a rule) a certain judgement as to the category or quality of the object. Rules in such a context are what join the properties with the category / quality of the object, e.g. someone smiling is taken as a property of someone being happy (the rule in question being that when someone is smiling this means that they are happy – one can also make this more normative, e.g. when someone hits a cat they are doing something cruel).

A context of systematisation is one where we make a claim or assertion about how two or more things are to be distinguished from each other (including potentially one being better or more appropriate than another), and in that kind of context we also tend to scan the environment for properties on the basis of which we can then make the two things distinguishable in a certain way while also being commensurable (i.e. where it makes sense to think of them as distinguishable in some specific way), e.g. I am interested in classifying two phenomena before me, and I look for properties that

make them commensurable (say both have two wheels, so I join that property up with them being a bike) and then distinguishable (I look further and see that whereas one has wheels of the same size, the other has a small front wheel, and so I refer to one bike in a certain way, and to the other bike in another way).

For the purposes of justification and systematisation, a person can treat what a certain rule demands as unchangeable, e.g. I will treat a rule that when someone is smiling they are happy, or that when something has two wheels (of whatever size) that something is a bike, as the way it must be. This 'taking', though, is active – it is not passive and exercised by regular default. Seeing that it is active helps us see that we are not determined to see it that way (i.e. just because I can treat someone smiling as meaning someone is happy does not mean I am determined, either by metaphysical or practical necessity, to *appreciate* someone as happy when, say, they are cringing). The contexts of justification and systematisation are two contexts where this active treating typically occurs, and these are the two contexts most often referred to in the Parts below. They may not be the only contexts in which we exercise this active treating, but they are central enough to do the job.

Once one sees this active treating as occurring within such contexts, one is well on the way to seeing that we do not always, and as it were by default, need to exercise such treating whenever we experience, act and interact. Our experiencing, acting, and interacting, does not involve the perpetual making of justifiable or systematisable claims or assertions. This is not easy to see if one remains within either a first-person or third-person perspective on social life, for adopting these perspectives sit well with prioritising the contexts of justification and systematisation. But as soon as one introduces a second-person perspective, and thus also sees persons as relating with others in a more tentative and responsive fashion – e.g. doing things with them, where this 'doing' requires both persons to be able to change the course of the other, rather than, say, figuring the other out from a distance – then one should also see that an alternative description of cognition in such relations is needed (precisely one that does not rely on justification and systematisation to do all, or even most, of the leg-work).

Parts V and VI will be showing, in much more detail, how distinctive the second-person perspective is (Part V), and then how distinctive is the picture of cognition it yields (Part VI). We will see that at the centre of the account of cognition from the second-person perspective will be a partnership between 'appreciation' and

'attentive silence'. This partnership is precisely the kind of cognition persons need to exercise when engaging in interaction understood from the second-person perspective. Appreciation plays a particularly important role in seeing the limits of the contexts of justification and systematisation. In other words, it is by making plausible the 'context of appreciation' that we see that we are not perpetually in, and thus cannot be described as if we were always in, a context of justification or systematisation. When we are in the context of appreciation, we experience someone being perhaps so-and-so in a way that is not mediated, personally or sub-personally, by properties presupposed as relevant, e.g. someone smiling, previously associated with that quality, e.g. someone being happy. In the context of appreciation, we can create the way in which it is relevant to appreciate someone as, say, happy – and we can create that way of appreciating someone as happy because of the way we experience the other, because of the way the other affects us or changes us, or because of the way we learn from the other in the course of interacting with them. This is not the moment to delve into details concerning appreciation – it is referred to, with examples, in Parts II to IV below, and is developed in detail in Chapter 24.

To summarise the above two Sections: conventions, norms and rules differ from each other. Conventions and norms are both closely related to regularity, in that both the behaviour that is the object of the convention or norm needs to be regularly exercised (though perhaps norms are less dependent on this being so than conventions are) and the attitude that expects (for conventions) or requires (for norms) that behaviour to be performed is also regularly exercised. Nevertheless, despite this similarity between conventions and norms, there are important differences too, for conventions tend to be addressed to the interests of persons whereas norms tend to be designed to defeat them, and, in addition, the normative attitude that characterises conventions tends to be weaker than that for norms. Rules, on the rational reconstruction offered here are not dependent in the way conventions and norms are on regularity. They are linguistic entities that are issued by certain authorities (recognised by others as being authorities). Nevertheless, they tend to be associated with order, i.e. they tend to be seen as relatively consistent sets that apply to a certain activity. What is key is that that order is not seen to be necessary, either metaphysically or practically, and that thus we avoid describing persons' abilities to participate in social life (as if it were a complex network of these orders) on the basis of some kind of internalisation (sometimes unconscious) of those rules. In short, we

should avoid regularising rules, characterising them, instead, as contingently useful resources that can, in certain contexts (typically justification and systematisation) be experienced as necessary (when the behaviour they demand is treated as unchangeable). From the second-person perspective we see the limitations of conventions, norms and rules as devices for either setting up the object of explanation or featuring in the explanation itself. The second-person perspective shows us that conventions, norms and rules are not determinants of appropriateness or inappropriateness, but rather contingently useful resources.

Part II. Necessity:
The First Element of the Governance View

Armed now with a basic grip of conventions, norms and rules – depicted as 'determinants of appropriateness and inappropriateness' on the governance view, and understood as contingently useful normative resources from the second-person perspective – we now need to begin focusing on the arguments of the governance view. In this Part, the focus is on various versions of the claim that determinants of appropriateness and inappropriateness are, in some sense to be further explained, necessary. This is the first element of the governance view. The second, that minds are governed by these determinants of appropriateness and inappropriateness, is the focus of Part III. The third element, that, at the social level, these determinants of appropriateness and inappropriateness are necessarily shared is the focus of Part IV.

Four versions of this first element of the governance view are discussed here: first, the argument from the very concept of appropriateness (Chapter 5); second, the argument from the possibility of error (Chapter 7); third, the constitutivist argument, which we shall examine in two different forms, one in terms of a theory about the nature of games (Chapter 8), and the second as a theory that treats language as if it were a game (Chapter 9); and fourth, the super norm argument, or the argument that postulates rules or norms that are inherent but invisible, *a priori* and implicit (but able to be made explicit) (Chapter 10).

In each case, the thesis seeks to replace necessity with contingency: what are thought of as determinants of appropriateness and inappropriateness that are necessary (in the different ways that necessity is cashed out) is better thought of (certainly so from the second-person perspective) as contingently useful normative resources that can sometimes, in certain contexts, be experienced as necessary.

One proviso should be made clear at the outset. References below will be made to (especially) rules and sometimes norms, but as this is done by specific theorists no claim is here made that the reconstruction of the terms provided in Part I maps onto these uses. Nevertheless, this does not render Part I redundant, for it is useful to get a basic, initial grip on the ways that these various terms have been used, and to clarify the approach the thesis takes to those concepts, when reconstructed.

Chapter 5. The Very Concept of Appropriateness

The focus in this Chapter is on that version of the first element of the governance view that asserts that determinants of appropriateness and inappropriateness are necessary because without such necessity the very concept of appropriateness would not make sense. What this Chapter seeks to show is that the viability or intelligibility of the concept of appropriateness need not rely on determinants of appropriateness or inappropriateness being necessary; instead, the concept of appropriateness itself includes the possibility of being revised in a way that reveals the limitations (the contingent applicability, or the contingent usefulness) of any normative resources that were previously thought to be useful in that domain.

One of the most sophisticated analyses of appropriateness – though using the term 'correctness', which is similar enough for present purposes[43] – has been provided by Judith Jarvis Thomson, in her book *Normativity* (2008). There, Thomson dedicates two chapters to 'correctness properties'. The first of these chapters deals with acts and the second with mental states. The chapters are rich with ideas, and some of them are best served by considering the arguments in the book as a whole. Analysing the book as a whole, however, is outside the scope of this thesis. Nevertheless, there is one distinction that Thomson introduces that we can look at in isolation form the rest of the arguments in the book, this distinction also being a useful insight into how the concept of appropriateness is used in the literature. Having discussed this distinction, we will then to the more general task of considering whether the very concept of appropriateness requires, by necessity, the presence of determinants of appropriateness (especially where this idea takes the form of a definition of appropriateness in terms of compliance with pre-existing standards).

A. Internal and External Correctness

Early on in the chapter on acts, Thomson makes an instructive distinction between 'internal' and 'external' correctness. The distinction is best understood with an example. Imagine that Jones and Smith are asked the following question: 'Who is

---

[43] There are some theorists who use the term 'appropriateness' in a broader sense – akin to the realm of the evaluative and thus the broadly normative – where the issue becomes whether appropriateness can be exhaustively explained in terms of 'correctness' (assumed to be the narrower usage). For present purposes, however, the distinction and this debate are not of central importance. The key issue is not whether the broadly normative can be explained in terms of the narrowly normative, but how the normative in general is understood.

Alice's brother?'. Imagine that Jones answers 'Bert is Alice's brother', and that Smith answers 'Bert are Alice's brother.' Imagine that both Jones and Smith are students in an English as a Second Language class. Imagine further that Bert really is Alice's brother. Internal correctness here refers to the correctness of 'how the agent did what he did' (Thomson 2008, 87); differently put, we can use the locution 'internal correctly' when 'we wish to assert the proposition that is true just in case X asserted that Bert is Alice's brother, and carried out the enterprise of doing so correctly' (Thomson 2008, 88; here, this applies to Jones, presumably because Smith violates a rule of grammar and thus does not 'carry out' the enterprise internally correctly). External correctness, on the other hand, refers to the propositional content of the assertion – what matters here is whether the content is true, not the manner in which it was carried out (Thomson 2008, 88; here, this applies to both Jones and Smith, for both identify Bert as Alice's brother). It follows that someone can be internally correct without being externally correct (imagine Brown saying 'Mark is Alice's brother'); externally correct without being internally correct (as in the case of Smith); and both internally and externally correct (as in the case of Jones).

It is interesting to see how Thomson uses this distinction. Thus, for example, she uses it to target the relationship of constitutive rules to the idea of correctness. It is common, she says, to assert that constitutive rules impose obligations. Take the constitutive rules of tennis or chess. These rules have been taken by some as providing an analogy to certain other practices, e.g. to the practice of asserting. Thus, Thomson shows that Timothy Williamson has argued that the 'Knowledge Rule' (which says 'One must: assert that p only if one knows that p') is like a constitutive rule of chess (Thomson 2008, 88-9; see Williamson 2000, chapter 11). In other words, we should think of asserting as if it were a game – in this case, governed by the Knowledge Rule. According to this line of thinking, 'A person who asserts that p when he does not know that p breaks a rule of asserting just as a person who moves a bishop horizontally while playing chess breaks a rule of chess' (Thomson 2008, 89). Thomson makes short shrift of this idea:

> Alas, for the idea, the rules of a game impose no obligation of any kind on the players. The rules of chess do not tell you what you are under an obligation to do. They do not tell you what you must, or even ought to do. Suppose you are playing chess, and it is your turn to move. You then learn that if you don't move your bishop horizontally, hundreds will die! Are you all the same under

an obligation to not do so? Must you, ought you not do so? That idea is just silly. (Thomson 2008, 89-90)

The point here can be explained by recourse to Thomson's distinction between internal and external correctness. The constitutive rules of chess determine, on Thomson's view, internal correctness (just like the rules of grammar determine the internal correctness of language use). They are the rules that dictate what 'counts as a legal move in chess' (Thomson 2008, 90; or a grammatically correct locution). But internal correctness is quite separate to external correctness, though in many circumstances to follow the rules that determine internal correctness will also coincide with what is externally correct. In this scenario, as Thomson presents it, external correctness might be taken to refer to when playing chess – when following the rules that determine internal correctness – is itself correct. Thomson's point, put in these terms, is that the rules of internal correctness do not guarantee their own applicability, and so it is too quick to argue that constitutive rules impose obligations.

Understood this way, the distinction is a useful one for it reminds us of a gap that we often do not see (arguably, the constitutivist strategy depends on us not seeing it), precisely because we will describe cases as ones where the rules that determine internal correctness are obviously applicable (i.e. our description will not question the applicability of the rules). Thus, we might describe a case as one in which persons are playing in a chess tournament, and if that is all that is said in the description, and if playing chess is defined as acting in accordance with such-and-such rules, then, given that no reason has been given for questioning the applicability of the those rules, one is not likely to question them (at which point we see no gap, for the description does not given us any reason to doubt the applicability of the rules). Thomson's splitting correctness into internal and external, then, is helpful for it guards against this forgetting that it is always possible to question the applicability of the rules. In this way, it puts the onus back on the constitutivist strategy, which is, by virtue of the distinction, effectively said to over-generalise from it usually being the case that we do not question the applicability of some set of rules to it always and necessarily being the case that their applicability is not questionable. In relation to Williamson's Knowledge Rule, all that Thomson would need to do, to put the onus back on Williamson, is to give an example of circumstances where we can avoid asserting, and indeed ought not assert, that p only if we know p – as when, for instance, we are

in a live TV show game in which we may need to assert the answers to questions we do not know.

Arguably, Gideon Rosen (2001) makes the same distinction, though in different terms. Rosen's distinction is between what he calls 'correct-making features' and 'correctness itself'; the former corresponds to internal correctness, and the latter to external correctness, in Thomson's discussion. For example, Rosen argues (thereby providing an example of correct-making features) that 'To play a Mozart's C major sonata just is to engage in performance that is correct only if certain notes are played in a correct order', or 'To be dancing the mambo just is in part for certain steps to count as correct' (Rosen 2001, 619; it is not explained what the 'in part' refers to in the second example). 'Correctness itself', on the other hand, is the applicability of the correct-making features – which, as said above, corresponds to Thomson's external correctness.

The distinction – whether in Thomson's or in Rosen's vocabulary – is a useful one for arguing against the constitutivist strategy (to which we will return later), but it also potentially yields a misleading concept of correctness (or appropriateness). This is because it makes it seem like there are domains where what is appropriate is necessarily determined by rules – it never being the case that, within that domain, one could appreciate something as appropriate without applying the rules of that domain, for the only other way in which one might find what is appropriate to do is by reference to a second-order consideration that renders those rules inapplicable (at which point one might say one has left that domain – one is no longer just playing chess but now determining lives by the movements of pieces). Further, it is also misleading because it makes it seem as if there is something like 'correctness itself' (or 'external correctness', in Thomson's terms), which is presented as having a final incontestable say on matters of correctness (we are asked to take the second-order evaluation of appropriateness, i.e. concerning the applicability of the first-order rules, as incontestable).

Put differently, the problem is two-fold: first, that what is appropriate is by default determined (at the first-order level), and the only way in which it can be said to not so determine what is appropriate is when the rules are not applicable (according to the second-order level); and second, that there is a second-order level ('external correctness', or 'correctness itself') that determines the applicability or non-applicability of the first-order rules. What needs to be shown is that there is a middle

path: between the distinction between internal and external correctness on the one hand, and, on the other hand, the constitutivist strategy. At stake is the very concept of appropriateness and whether rules, norms etc are necessary for that concept to make sense.

Let us try to put this another way: if our only argument against the constitutivist strategy is one that relies on it being the case that things are appropriate only when the rules that ordinarily determine appropriateness are non-applicable (by virtue of 'correctness itself' determining that non-applicability at the second-order level), then we will only further entrench, and not make it easier to move away from, the idea that appropriateness is necessarily dependent on the application of rules. So we must find a way of showing that it can never be enough – that it is never necessarily enough – to rely on the rules (at the first-order level) for appreciating appropriateness, but that that not being enough does not mean that we must rely on either a second-order set of rules or some notion of 'correctness itself' (exercised at the second-order level).

It helps to switch the matter round: why would we want to say that the only way in which we can evaluate whether something is appropriate to be done is by saying that we are making that evaluation on the basis of some rule? In the case of appealing to 'correctness itself', it may not be that we are appealing to a rule, but then we do seem to be appealing to some implicit status that our evaluation of the appropriateness of something is being deemed to have, i.e. we are describing it in such a way that it does not appear to be contestable whether it is correct (the only facts in the description are ones that support the second-order evaluation, e.g. that it is inappropriate to move the bishop, and thus follow that rule for moving a bishop, if hundreds will die – here one can ask, by way of challenging the description: but what if hundreds more will die as a result of us not moving the bishop?). Appropriateness on the view we need to resist is never tentative: it is always certain. But why should appropriateness be always determined, either by a rule or by some allegedly implicit status created by the description of the case?

Again, as this is not easy, let us try to articulate it one more time: if one limits the idea of appropriateness to either the application of rules at some first-order level or to the non-applicability of those rules as a result of the exercise of some inherently appropriate judgement at the second-order level, then one's tale of appropriateness is of two determinisms, i.e. the determinism of first-order rules and the determinism of

second-order status. To avoid this, one must give an example where someone can appreciate something to be appropriate without relying on a rule and without that appreciation being guaranteed a certain status (and thus magically, as it were, said to be exercised at a higher-order level). The matter is difficult because the second-order level is useful to appeal to – but only when one assumes that there is some first-order where what is appropriate is determined by rules that are by default applicable unless shown otherwise.

To see that we need not depend on the distinction between internal and external correctness, and thus that we can loosen the notion of appropriateness from determinants attached by default and yet not accept constitutivism, return to one of Rosen's examples: 'To be dancing the mambo', said Rosen, 'just is in part for certain steps to count as correct.' One unsatisfactory way to question the example would be to say that that it can only be appropriate not to make those steps in the dance when making the steps in a different way would be appropriate because of the non-applicability of the rules requiring those notes and those steps. Again, this is unsatisfactory because it to say this is to replace one determinism with another (replacing a first-order determinism with a second-order one). Instead, we need to say something like the following: there may be different ways of dancing the Mambo, and these different ways, though they will not be 'automatically appropriate because they are in accordance with the default rules' may nevertheless be appreciated and appreciable as appropriate. In other words, it is only when one commits oneself to saying that there is but one way, the only one and the only one that will ever be, to dance the Mambo that one is obliged to accept a concept of appropriateness that is always determined (either at the first- or the second-order level). In other words, thinking of the concept of appropriateness as necessitating determinants requires assuming that there are set ways, exclusive ways, of doing things, as if labels such as 'the Mambo dance' were metaphysically fixed – as if there simply is no other way of doing them (for if one does them in a different way, one is no longer doing them, but doing something else).

Are the names of things attached so tightly to specific ways of doing them? Notice that here there are two different things one could say: first, one could say that there are two categories of things, those that must be done and can only be done in some specific way, and those that can be done in a variety of ways; or, second, that all things could be done in a variety of ways, for there is no authority capable of such a

power that would control what name can attach to what ways of doing things. The first helps to make room for the second, but it is only in that sense that it is helpful – for in itself, the first statement is misleading. Thus, it may be helpful to say that highschool dances are less regulated – there are less specific ways of doing them – then, say, the ceremonial dances performed at the borders between India and Pakistan. But if one took this to mean that there was only one way, and there could only ever be one way, of dancing the 'Border dance', then one would be over-generalising the lesson to be learnt from the comparison. What would it mean for there to only be, and only ever be, one way of dancing the 'Border dance'? For a start, it would mean that it could never change – that it would be timeless, a timeless form as it were to which one either conformed or did not. Second, it would mean there would have to some authority that polices what could be called a 'Border dance', which connected it mechanically to certain very specific ways of dancing. Third, it would mean that it could be – indeed would need to be – danced by machines programmed to so dance it, where these machines would just repeat exactly the same moves each time and go on so repeating them until and unless we programmed them otherwise. Are any of these plausible commitments for an understanding how humans engage in ceremonial dances – or any other activities?

Imagine the following possibility: someone dances a Border dance and as they do so, they are bothered by a bee; being bothered by a bee, they make moves that they might otherwise not have made; but the person making the moves is one of the most respected dancers in the group; and so, the other dancers come to adopt some of the moves of that person; and so those moves come to be appreciated as part of the Border dance. This example helps a little to see how and why it is impossible to completely control (for no authority has such a power to control) what could ever count as a 'Border dance', but it is still somewhat unsatisfactory for 1) it suggests a default that remains the same until it changes; and 2) it makes it seem like it was inevitable that the moves introduced by the respected dancer would become part of the dance (to avoid this second consequence, one would need to say that those moves may or may not have come to be appreciated part of the dance).

To improve on this, let us say the following: what makes the Border dance a human dance, and not a machine one, is that it is never the same; it is always subtly different. There is no set form to a Border dance, but we can (especially as observers) treat it as if it had a form that never changes or could never change (perhaps for the

purpose of describing it in comparison, say, to some other ceremonial dance). When persons dance the Border dance, and especially when these persons are the ones who always perform it (imagine a troupe who specialises in it), they are never quite repeating the exact same moves. The fact that the dance is not mechanically reproduced or reproducible is what allows it to change, and change precisely in ways that are not controlled (or necessarily introduced intentionally) by anyone. As soon as one sees this, one also sees that one can appreciate a Border dance as appropriate without it reproducing the same moves, and thus without classifying it mechanically under some rule-structured category; this does not mean appreciating as appropriate 'despite' it not reproducing those moves (that would take us back to the default-unless-given-reason-otherwise model) – it means appreciating in a way that nobody else may have thought it relevant to appreciate it (imagine, for instance, that one came to appreciate, for the first time, the way these dancers move their heads, whereas what was always described were only their feet movements). Of course, this also means that one's appreciation of appropriateness (in the above example of the head movements) is more tentative – it is not guaranteed to be accepted / acceptable by others, for its status is not safeguarded, i.e. it is precisely not presented as being made on the basis of rules or norms others also share or on some allegedly absolutely objective basis (of 'correctness itself').

The general point is this: the argument that the very concept of appropriateness is one that necessarily depends on the application of determinants (of some set rules that specify what ought to be done) only works when one assumes that there is a set way – and only that set way – of doing things (performing the Border dance), such that there is also only one way of appreciating them (appreciating the dance as a Border dance). It is as if one assumes that any one description of how the Border dance ought to be done is exhaustive: there is absolutely nothing else that could ever be relevant to performing / appreciating the Border dance (similarly with the Mozart sonata: it would have to be the case that there could never be anything added, no extra notation for instance, no alternative finger movements). But this is an implausible assumption – it would mean that ways of doing things, as described, are so controlled that there is an authority that programmes machines to reproduce those ways of doing them, such that those things would only change when the machines were re-programmed to do them a different way (the authority would thereby create a new name for that form, or delete the previous form and attach the old name to this

new form; either way, the authority would always be in control of what and how things can change).

Things are very different if we avoid giving acts of appreciating appropriateness an implicit status as accepted or acceptable (as justified or justifiable), and thus avoid making it necessarily dependent on rules or norms (that would give it that status): now it becomes possible for persons to dance not only highschool dances, but also Border dances, in non-mechanical, non-automatic and, in short, non-deterministic ways. Making room for the tentativeness of appreciation of appropriateness goes hand in hand with an underlying indeterminacy – of a mix of the regular and irregular, of the predictable and unpredictable, of what has been understood so far to be relevant and what could emerge as relevant – of human activities, and in doing so points to the inevitable limitations (the contingent usefulness) of any particular specification of how they ought to be done. This still leaves it open for some persons to be particularly committed to certain specifications of how things ought to be done, but it no longer makes it necessary for all persons to be like this all (or even most) of the time.

What has been attempted in the above paragraphs is a loosening of the concept of appropriateness, including showing how narrowing it tends to be made under certain pressures for certain purposes. We proceeded dialectically, at first setting up a distinction made in the literature between internal and external correctness (though also going under different labels, e.g. first-order versus second-order appropriateness), and then going on to show that although this helped to critique the constitutivist strategy, it in effect entrenched a certain understanding of appropriateness that we need to avoid. We then unravelled this understanding, and made some first small steps towards an alternative (towards a more tentative appreciation of appropriateness that is not automatically guaranteed any sort of status). In so doing, it is hoped that we have a come a little closer to peeling off attachment to the necessity for determinants of appropriateness.

From the second-person perspective, the concept of appropriateness is one that includes within it the inherent revisability that is part of it. Appropriateness ought not to be thought of, in the first instance, as a closed domain – remaining closed until we choose to revise it. It is not, at bottom, an order that is defeasible. Instead, in the first instance and at bottom, it is a mixture of the expected and the unexpected, of the known and the unknown, of the recognised and the as-yet unrecognised, but we can,

for certain purposes, at certain times, treat something as constituting an order, as being closed normatively. As we shall see in later Chapters, when we interact with another we can treat as unchangeable certain requirements of those conventions, norms or rules we deem to be useful; but by no means is this the only or even the most important kind of ability we need to participate in social life. Instead, what we see we need from the second-person perspective is a much more flexible, more tentative approach to what might be appropriate. The point above is just to show that the very concept of appropriateness need not be characterised as one that requires it to be closed, to be exhaustively populated, at the outset, by determinants of appropriateness and inappropriateness.

B. *Appropriateness as Compliance*

Before we leave this Chapter, we need to say a few words about a popular strategy, namely the explanation of appropriateness via the notion of compliance.[44] Having dealt with the distinction between internal and external correctness, the appeal to compliance is easier to defeat, for the drawing of a first-order and second-order split (both of which are imagined as determinative) is the most sophisticated way of arguing that the very concept of appropriateness requires, by necessity, determinants of appropriateness. The appeal to compliance, by contrast, is limited to the first-level, i.e. it sets up a domain or environment said to be necessarily governed by rules such that the only way in which something can be appropriate in that domain is by compliance with those rules. It is important, in arguing against this appeal to compliance, not to draw on examples that would create the need for a second-order level of rules that determine the applicability of the first-level rules – that, as we have seen above, is not the way to argue against the necessity of determinants of appropriateness (for it just replaces one determinism with another). Instead, we must show that there are infinitely many possible situations where the rules, as we know them to apply to a certain domain, are simply silent and irrelevant to appreciating something as appropriate.

We can work our way towards the kind of example we need by first considering the following case: there is a rule in chess that says that when one

---

[44] That it is popular can be ascertained by considering any dictionary's definition of appropriateness or correctness, where reference is quickly made to 'conformity' or 'compliance' with standards (or, in general, determinants of appropriateness).

promotes a pawn to the end rank, one can get any piece one likes, except for another King (or, of course, allowing the pawn to remain a pawn: one must actually get another piece, but just not a King). Amateur players often misunderstand this rule, mistakenly thinking that the rule is that when one promotes a pawn to the end rank one must get a Queen. Imagine Sally is playing chess with her brother, Simon. Sally promotes a pawn and gets a Queen. But Sally is – let us say, for argument's sake – following a rule according to which one must get a Queen. There is a sense in which Sally's promotion to a Queen 'complies' with the rule of chess, but there is also a sense in which there is something unsatisfactory about Sally's behaviour, i.e. there is a sense in which we can say her behaviour is not quite appropriate (for – again, let us use this way of putting things for the moment – it does not quite follow the right rule). If it is plausible to say there is something incorrect about Sally's behaviour, then it also seems plausible to say that Sally's behaviour is inappropriate, and yet compliant (if someone was observing Sally's behaviour, and was mindful of that rule of chess, they would have no grounds – unless, perhaps, they interviewed Sally – for asserting that Sally did not comply with the rule).

This example can help us, initially, to introduce a gap between appropriateness and compliance, but it is also unsatisfactory. The basic problem, from the perspective of this thesis, is that someone could say that it appeals to what it wants to question, namely that appropriateness is determined by some rule – in this case, the correct rule (Sally's behaviour is inappropriate, and thus non-compliant, with the rule, and that is why compliance in this case is not enough to establish appropriateness). Given this possible reply, perhaps the best way to use the example for our purposes is to say that even compliance, when appealed to, requires some reference to what may be the relevant way of complying, the point being that the rules themselves do not tell us what might be (or, better, what we might appreciate as being) relevant to complying with a rule (in the above case, we need to decide – and nothing decides for us, certainly nothing in advance of us so deciding – whether Sally's thinking that she must get a Queen is sufficient for compliance). This is helpful, for it points to the limits of the very appeal to compliance, but we can, and need to go further, for the above case still focuses on the applicability and scope of the rules.

Imagine, then, the following case: in the course of playing Simon, Sally, upon promoting her pawn to a Queen, says 'I promote!'. There is no rule according to which what Sally does is appropriate (or indeed inappropriate). The rules are silent,

and in a way irrelevant. One could, certainly, draw an analogy with saying 'Check' or 'Checkmate', and make a claim, on the basis of this analogy, that Sally's announcing promotion is in line with (it is systematic with) the other rules (the ones that allow for saying 'Check' and 'Checkmate'). Or, one could claim that Sally so announcing is justified, for, say, it signals to Simon something he should be given a chance to notice (perhaps because, if they were playing quickly, he would otherwise miss it). Or, perhaps one could say it is justifiable because Sally is teaching Simon to play (including what it means to promote), and so it makes sense, and is thereby justifiable, for Sally to announce promotion (here, appropriateness follows intelligibility, where intelligibility is understood as narrowing down the intelligible means for pursuing certain ends). One could certainly do either of those two things, thereby describing the scene under the aegis of the contexts of justification and systematisation, but if one did so – if one hurried to explain the case this way – one would be neglecting to see that the first time Sally says 'I promote', an appreciation may occur (an appreciation is not necessarily required – Simon could ignore Sally's announcement, and Sally might never go on to repeat it – but it may occur).

This potential appreciation is exercised without recourse to a rule: Sally, having said it, or Simon, faced with Sally saying it, or someone observing them playing, could appreciate it as either appropriate or not, and if appropriate, could appreciate it as appropriate in a way that they need not have appreciated (or claimed) anyone to have done something appropriate when playing chess before. Someone could, for instance, appreciate Sally as announcing promotion as appropriate and in so appreciating her as appropriate create a ground that could become relevant for playing chess (in this house, say): e.g. that, when learning to play chess, it is a good idea (and, say, in this house a requirement), to make an announcement whenever one plays a move that yields a big advantage (e.g. one might go on, on the back of this, to announce whenever one takes another piece of higher value: 'Value Gained!').

The general point is this: appeal to compliance with pre-existing rules assumes that for the purposes of evaluating the appropriateness of something within that domain (to which the rules are assumed to be applicable) all that is relevant, and could be relevant, to consider is the applicability or non-applicability of the rules, i.e. whether the rules are complied with or not. However, as we have seen above, in the first place, the evaluation of compliance is itself dependent on a certain understanding of what the rules (assumed to be applicable), require (in the first example above, the

person who says that Sally complies assumes that for the rule to be complied with, it is sufficient if Sally promotes to a Queen, even if she thinks that she must get a Queen – the point being that this commitment to sufficiency can be challenged). And, further, in the second place, explaining appropriateness by compliance with the rules assumes that the rules will exhaust what may be relevant to any appreciation of appropriateness. The point, then, is that explaining appropriateness by compliance assumes, and is not warranted in assuming, that relevance is exhausted in advance. Put differently, explaining appropriateness by compliance positions us perpetually passive vis-à-vis that which can emerge as normatively relevant.

It may be that the tendency to want to explain appropriateness by compliance comes from the desire to avoid a scenario in which someone can say anything at all – as if randomly asserting that something is appropriate for any reason whatsoever. On the back of this worry (of arbitrariness) one insists that one can only ever find appropriate that which has been previously granted legitimacy to be so found – as if one would need to, by necessity, whenever appreciating the appropriateness of something, appeal to grounds that others ought to (e.g. because it was previously agreed to) accept as legitimate grounds for any appreciation of appropriateness. This is an understandable worry, but the problem is that it sets up two extremes: either absolutely conformity or absolute arbitrariness. It is true that absolute arbitrariness needs to be avoided, and it can be (e.g. by saying that if one wants to 'claim' or 'assert' appropriateness, then one must appeal to grounds likely to be found warrantable by others, or indeed oneself). The point being made here is that there is something in the middle, between exhaustive conformity with pre-existing rules and someone's arbitrary say-so, and this something in the middle is the tentative exercise of appreciation of appropriateness. This tentative exercise of appreciation is precisely something that can (though it need not) result in the adoption of some ground for future treating certain behaviour as required (and unchangeably required, e.g. the future treating, in this house, of announcing some advantage gained or risk to one's opponent, as required and unchangeably required). It is, in other words, something that positions persons as active vis-à-vis what may be relevant: not perpetually and arbitrarily active (in some blindly autonomous way, where our say-so is trumps), but active in a way that is subject to being responded to, in unpredictable ways, by others. It may, for instance, be appreciated as an irrelevant ground by someone else – the point is that this cannot be described, by the theorist, as somehow inevitably

resolvable in some way, for when it is so described the theorist assumes and imposes what is and what is not relevant; the point, in other words, is that relevance is indeterminate for the parties, but may be treated by them, actively, as being determined.

We will be saying more about appreciation, including appreciation something as appropriate or inappropriate, later in this thesis (Chapter 26); for the moment, what is important is to see how we need to make room for it, and thus why it is inadequate to seek to explain appropriateness via compliance (or, as in the preceding Section, by appealing to two determinisms: first-order and second-order, thereby making it seem as if there will be cases where a first-order determinism will be sufficient). Let us now proceed to examine how and why the possibility of error – and thus inappropriateness – is likewise inadequately understood when it is said to require, and be determined by, determinants of inappropriateness.

Chapter 6. The Possibility of Error

The possibility of error argument – the second way in which the first element (necessity) of the governance view can be cashed out – is an important and often-encountered one in the literature on normativity. In its most friendly form, the argument simply states that it is a necessary condition for the very existence of the realm of the normative that something can go wrong – that we can think that something has gone wrong (either when reacting to something we ourselves have done, or something others have done). This is a 'friendly form' from the perspective of this thesis because it does not, in that general formulation, make the possibility of error necessarily dependent on the application of rules or norms (or any other determinants of, in this case, inappropriateness). But, as we shall see in a moment, the argument often takes on a form that gives it pride of place in the arsenal of the governance view, i.e. it is argued, or quite often simply assumed, that in order for something to be mistaken there must exist phenomena, i.e. criteria for (or determinants of) what is mistaken, or in error, or wrong, or inappropriate. Read this way, the argument becomes another way of arguing for the necessity of determinants of appropriateness and inappropriateness. What shall be argued here, to the contrary of this stronger (determinant-dependent) reading of the argument from the possibility of error, is that it is entirely possible to give sense to the idea of errors, mistakes or inappropriateness without it being necessary for that error, etc, to be a violation of a determinant of what is inappropriate.

Our procedure here will be as follows: we shall begin by looking at a recent, detailed examination by Douglas Lavin (2004) of the possibility of error argument, and see how the notion of error is understood in it. Having done so, and having a better grip of the stronger reading of the argument, we shall then consider a paper by Paul Roth (2003) that can help us move away from this stronger reading, and retain the possibility of error but without positing and necessitating determinants of inappropriateness. As with all the Chapters in this Part, our aim throughout is to resist a reading of conventions, norms and rules that would render them necessary – showing, instead, that they are contingently useful resources – while allowing them to be experienced as necessary for certain purposes in certain contexts.

A. Error and Agency

At the beginning of his paper, Lavin cites several quotes that show, in his terms, 'the deep connection between reason and the possibility of error' (Lavin 2004, 424). It is instructive that the first of these is: 'For a creature to be correctly said to have a rule, it is necessary that it should be able to break the rule' (Lavin 2004, 424; the quote is from Bennett 1989, 17). This already should indicate the difficulty we shall have in extricating the possibility of error argument from an argument based exclusively on the possibility of breaking rules. Indeed, Lavin's paper is set up precisely by this dialectic – where to act appropriately (or rationally) is to act in accordance with the rules that govern rationality, and where to act inappropriately is to break those rules. Thus, his opening paragraph is as follows:

> Insistence on the link between reason and the possibility of error is just insistence on the normative character of reasoning. On any account, reasoning is activity governed or guided by norms, rules, standards or principles, and so the very idea of this activity must contain the distinction between correct and incorrect application of them. Indeed, to most anyone who has thought about what a principle is and what it is for an individual to be responsible or subject to such a thing, it will seem that: a reasoner is subject to a principle only if the reasoner can go wrong with respect to it. And on this basis, most philosophers have accepted as a condition of adequacy on any account of being subject or responsible to a principle that it be able to distinguish between a reasoner's correct and incorrect application of the principle, between her faithfulness and unfaithfulness to it. (Lavin 2004, 425)

Notice how closely aligned the following two theses are assumed to be: first, the thesis of the possibility of error; and second, the thesis of the possibility of not following (or not being faithful to) principles. Further, Lavin suggests it would be wildly implausible to disagree that reasoning is an activity governed or guided by norms, rules, etc.[45]

Lavin, then, correlates the possibility of error argument with the possibility of violating a principle. He explicitly refers to the 'error constraint' as the possibility of an agent going wrong with respect to a principle (Lavin 2004, 425). In fact, as his own analysis very early on in the paper indicates, the possibility of error runs deeper

---

[45] As an aside we should note that the phrase 'governed by' or 'guided by' can be used in ways more or less sympathetic to the governance view. If, for instance, by 'governed by' one means that there are normative resources that are contingently useful for the participation in and understanding of some activity, one need not be embracing the governance view. The point is rather whether, in using that phrase, one means that there are determinants of appropriateness and inappropriateness that necessarily govern that activity (or domain, etc).

than the (general formulation of the) error constraint suggests, for it is involved in the very definition of the idea of a 'principle' and of what it means to be 'subject to it' (indeed, one might say the concept of agency also falls under its jurisdiction). Thus, for example, a 'putative principle formulable as "Do A or don't do A" or "You must either do A or else not do A" isn't something which can agent can violate' (Lavin 2004, 426) and for this reason cannot be a principle. 'The logical interpretation of the error constraint', then, is 'really…about the principle' (Lavin 2004, 426-7), about what the principle must be like to count as a principle. The way that Lavin formulates the logical interpretation, though, is somewhat misleading, for he says that on this logical interpretation 'an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle' (Lavin 2004, 426). This is somewhat misleading because it contains the term 'principle', and thus uses something in its own definition that it seeks to define. In any event, Lavin's point is that the logical interpretation of the error constraint is not interested in 'facts about the agent' (as is what he calls the 'imperatival interpretation' of the error constraint). On the imperatival reading, one looks not to the content of the principle (to determine whether it counts as a principle, i.e. whether it can be violated), but rather to certain facts about the agent, or to 'the coupling of the agent to principle, or again…[to] the specific nature of an agent's relation to the principle' (Lavin 2004, 427). To be able go wrong on this imperatival interpretation means 'that it is possible for the agent to act out of accord with the relevant principle, perhaps because of *akrasia*, ignorance, vice, or some other feature of the agent' (Lavin 2004, 427). The full specification is that on the imperatival interpretation 'an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle and it is possible for the agent to do it' (Lavin 2004, 427).

The split between the two interpretations is puzzling. This is because it looks as if the imperatival interpretation is dependent on the logical one, at least in the first instance – and this is especially so if, as Lavin suggests, the logical interpretation is about the form of the principle. After all, the imperatival interpretation contains the term 'principle', and so presumably it applies only to principles that count as principles – which, as noted above, it is the job of the logical interpretation to define (or at the very least to exclude certain kinds of formulations). Ironing out this puzzling split, one could try to reformulate the imperatival formulation in the

following way: 'an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate something that can be violated and it is possible for the agent to do it.' This seems to be an improvement, until one tries to articulate what 'something that can be violated' means, and one ends up where one started.

Lavin has serious doubts about the plausibility of the error constraint on the imperatival interpretation (and we shall come back to this in a moment), but he harbours no such doubts about the logical interpretation. In fact, he explicitly says that the logical interpretation is 'obviously true' (Lavin 2004, 427). This is a very strong claim, and indeed can be shown to be hardly so obvious. Go back to the example that seems clearest: 'Do A or don't do A.' One can imagine a circumstance where this could be something that could be violated, e.g. it could be completely ignored, whereas it is designed to be taken into account such that one ought either to do A or not do A (and not, for instance, do C or B etc). In other words, this could be an instruction to an agent to not consider doing anything else but doing A or not doing A – what is important is that the agent does one or the other. In this sense, violating the instruction would be, for instance, doing B. Even if the formulation were to be changed to 'Do A *and* don't do A', then one could still say that this instruction asks one to first do A, and then to not do A (so, say, first smile, and then not smile). One could then try to add, 'Do A and don't do A at exactly the same time', but one could of course make this work as well as an instruction (e.g. one records oneself as doing A, and then records oneself as not doing A, and plays both tapes at the same time on different screens), and so on. The point here is that it is not all clear that there is any logical interpretation that, by itself as it were, determines whether something is or is not a principle.

Lavin appears to recognise that the interpretation of the principle can be a factor in determining whether it can be violated (he recognises some difficulties of this kind with non-principles at Lavin 2004, 430-1), but he thinks that his 'Do A or don't do A' 'wears its emptiness on its sleeve and so only appears in real life as the expression of helplessness or frustration. "Keep drinking or don't. I don't care"' (Lavin 2004, 431) – which is a possible way of interpreting the phrase, but certainly not somehow more obvious than the reading suggested above, namely where what matters is that one does A or doesn't do A, as long as one does not do B.

Let us return to the imperatival interpretation, for this is one that is the focus of Lavin's paper. Recall that the focus here is on something going wrong with the agent, or better, something going wrong in the agent that damages the relationship between the agent and the principle. It is worth pausing for a moment to say that Lavin also presents his paper not merely as a paper about the error constraint (on his interpretation of it), but also about the conditions of agency. This is because Lavin thinks that agency is about being subject to principles. As he puts it, 'I am assuming that being an agent, that is, having the capacity to employ concepts in the service of action, has an essential normative dimension which I have expressed as being subject to principles' (Lavin 2004, 427). Given this assumption, he adds, 'we might also characterise the main thread of this essay as an inquiry into whether liability to error is a condition of agency or will' (Lavin 2004, 427). There is one important point to make about this. If Lavin does show that the error constraint (on the imperatival interpretation) is problematic, this does not necessarily show that the error constraint is not a constraint on our understanding of agency or normativity. What it could be said to show instead is that: 1) what is mistaken is Lavin's interpretation of the error constraint as being a matter of being able to violate principles; and 2) what is also mistaken is Lavin's assumption that being an agent (or the normative dimension of agency) involves being subject to principles. As an aside here, it is worth observing how much the characterisation of the 'normative dimension' is linked to how one understands human cognition (notice, in this respect, that Lavin admits to an understanding of it that emphasises how action is determined by the application of concepts).

In elaborating on the imperatival interpretation, Lavin focuses on what he takes to be Christine Korsgaard's use of it (in her 'The Normativity of Instrumental Reason', 1997; reprinted as chapter 1 of Korsgaard 2008). Lavin discusses how Korsgaard's argument against the Humean idea that agents are only instrumentally rational is dependent on the notion that merely instrumental agents cannot violate the instrumental principle. Lavin dismisses the possibility that Korsgaard's understanding of violation is simply a matter of the logical form of the principle, and suggests that Korsgaard's focus must be on some defect in the agent. In trying to get at the exact kind of defect, Lavin introduces a further distinction between two kinds of ways in which the imperatival interpretation could work: the first, a weak interpretation, does not pay attention to why the agent violates the principle; the second, the strong

interpretation, 'adds a further condition, one having to do with the source or origin of the error' (Lavin 2004, 436). The point here is that Korsgaard, on Lavin's account, needs to be committed to the strong interpretation, for she needs more than violations that come from, say, mere forgetting or mere ignorance (Lavin 2004, 435). She needs, Lavin says, the right kind of error – what he calls a 'genuinely practical error' (Lavin 2004, 436). According to Lavin, the right kind of error that Korsgaard has in mind is 'the sort of violation, error or going wrong that…is exemplified in cases in which "people's terror, idleness, shyness or depression is making them irrational and weak-willed' (Lavin 2004, 436). The point, with respect to the above case against instrumental reason, is that for Korsgaard instrumentality is not all that that there can be to reason because the instrumental principle is not one that an agent can make a genuinely practical error about. Reason must be something that we can be unresponsive to, but unresponsive not merely because of mere forgetting or ignorance, but unresponsive in a way where we see there is something that ought to be done but nevertheless do not do it (because, for instance, of terror, idleness, shyness or depression). For Korsgaard – on Lavin's reconstruction – the instrumental principle is not the kind of thing we can be unresponsive to in that way.

One relevant matter to consider in this context is whether there is a distinction that can be made here between mere forgetting and ignorance on the one hand, and, on the other, terror, idleness, shyness or depression. Further, if there is a distinction, is this a distinction that can be attributed to different sources of defect? And, if so, are these defects that are somehow necessarily tied to the violation of principles? Indeed, more strongly, could not these defects show that what we must be talking about here is not violation of principles, but something else entirely – something indeed important to agency, but not something that has anything necessarily to do with norms, rules, etc? The answer here is that indeed, there is an important distinction to be made between seeing that something ought to be done and not doing it, and not seeing that something ought to be done at all. The second is the defect that is more likely attributable to forgetting and ignorance, and the former more likely to be a cause of terror, idleness, shyness or depression. In that sense, the distinction is a good one and it can be attributed to different sources of defect. But it is also true that it is not necessarily the case that the distinction, and the presence of one defect (the one attributable to terror, etc) rather than the other, tells us something about the need for principles, or tells us that it is the way in which we relate to principles that allows us

to make the distinction between different kinds of defect. After all, what ought to be done (this being what I see, but do not act on) need not be a norm or a rule: I can see that Sally is sick, and this gives me reason to stay by the bedside with her and try to help her to forget her pain. I then stay and do precisely that. This entire process is not one that is necessarily mediated by my recognition of a rule (or my responsiveness to a norm). The reason in question can be a particular reason, and my noticing it and acting as I do can be something I do for the first time (I need not, for instance, have acquired a general model of being kind to persons who are sick, such that I apply that general normative knowledge to the present case). Further, although it could be said that my failing to see that Sally is sick, and also failing to see that Sally being sick gives me a reason to stay and comfort her, are all errors, it would be stretching things to say that these errors are necessarily (and thus only) explicable as violations of norms or rules (is there really a norm or rule that requires us to notice that persons are sick, or to comfort those whom we notice are sick?). What is happening here is that we may be missing what we may want to call 'defects in agency' if we focus only on the relevant way in which persons can violate norms or rules (and thus when we think that the only way in which they can do something that can be called an error or mistake is when they fail to be determined by a rule or norm that we, on the back of the description of the case, assert they ought to have been).

Lavin is critical of imperativalism (i.e. the imperatival definition of the error constraint), and relies instead on the logical interpretation. But his argument against the imperatival interpretation, and his reliance on the logical interpretation, are both developed in the context of what it means to violate principles – not what it might mean to have a defect or make a mistake more generally. And yet, his engagement with the imperatival interpretation suggests – as in the preceding paragraph – that there is something that is being missed here, or misdirected, as a result of being driven by a concern with how principles can be violated (as opposed to errors or mistakes made in general). What the various attempts to fix the imperatival interpretation are revealing is that a focus on how persons may violate principles does not go deep enough into the kinds of abilities (or absence of these abilities) that persons need to interact more or less successfully with other persons. None of this is to say that it is not important, and does not sometimes help, to be able to represent to oneself, to be able to commit to, to be able to be guided by, and so on, normative resources – all it says is that a focus on the possibility of error read as necessarily a matter of violating

a norm or rule is much too narrow a focus, i.e. one that misses out on certain other vital abilities that are not necessarily matters of being sensitive to rules or norms (e.g. such as seeing that Sally might be sick, and that she might need comforting). It is the risk and danger of missing these abilities that is part of the reason why it is so important not to restrict either the concept of appropriateness, or the possibility of error, to acting in accordance with / following or violating alleged determinants of appropriateness and inappropriateness. Another reason, of course, is simply our understanding of these concepts – our understanding suffers when we restrict them to the roles allegedly played by determinants.

Let us look a bit further at Lavin's discussion of the imperatival interpretation (on its strong version) of the error constraint, and see how else this discussion reveals to us the limits of understanding the error constraint in terms of an agent's relationship to principles (or rules / norms, which Lavin uses interchangeably in the paper).

One of the effects, as is clear in Lavin's account, of limiting the error constraint to the violation of principles is that – as is particularly obvious in Lavin's discussion of the imperatival interpretation – great emphasis is placed on the thought process by which the relationship of the agent to the principle must proceed. As Lavin says (interpreting Korsgaard), 'the efficaciousness of a principle must be mediated by the thought of what follows it' (Lavin 2004, 444). It follows, for Lavin, that 'the relevance of a principle to what happens in the world is mediated by the conceptual activity of that which is under it. The contrast is, of course, with the rather unmediated grip that a physical law has on what is, in some other sense, subject to it' (Lavin 2004, 444). We are already here narrowing human cognition to the self-conscious and reflective (Lavin 2004, 445) exercise of conceptual activity, and this should raise the following question: ought we to characterise the error constraint, and by extension the normative domain, in such a way (i.e. as a matter of the nature and function of principles) such that it makes us exclude all other features of human cognition and action that is not self-conscious, reflective and conceptual? Ought this effect already not put us on guard against such an understanding of the error constraint?

What is interesting in Lavin's paper is that the further he digs into the kind of relation needed for the imperatival interpretation to work, the more he gets into territory that is best understood not via the rubric of persons' relationships to

principles, but something else, e.g. persons' relationships to other persons. Thus, drawing on certain passages from Kant, Lavin speaks of the requisite thought process as one of indifference: what we need, so as to be subject to principles in the right way, is being capable of being indifferent to those principles (Lavin 2004, 448). But here we might ask: if we surmise (as we very well might) that the possibility of indifference is important to practical agency, why ought we to narrow our investigation in such a way as to make principles the object of that indifference? Why cannot the object of indifference be, for instance, the feelings of others? When Mark sees that Sally might be sick, but does not stay and comfort her, is he being (necessarily) indifferent to some principle, or is he possibly (perhaps even more likely) revealing indifference to how Sally feels or might feel should someone help her? In order to stay and comfort Sally, is it really the case that 1) Mark needs to have acquired knowledge of some relevant principle; 2) see that the principle is applicable and how it might apply in these circumstances; and 3) overcome the potential indifference (or any other relevant defect) and do what he understands the principle to require? Or is at the very least possible, if not more likely, that 1) Mark needs to notice that Sally is sick and in pain; 2) care about her feelings; 3) care sufficiently to stay and comfort her in the hope that doing so might alleviate her pain? Nothing in the latter process needs to be mediated by a principle – it only needs to be so if we think (mistakenly) that we can only come to do what we feel we ought to do when mediated by some self-conscious, reflective and conceptual activity (which is precisely what we come to think when we focus our efforts on persons' relations to principles).

For present purposes, we have gone far enough into Lavin's paper.[46] It should be clarified that the argument above is not one that targets an investigation into persons' relationships to principles *per se*. The point, instead, in the context of this Chapter, is how a narrow understanding of the error constraint – as a matter of violable / violating principles – has a pernicious effect on our seeing and understanding the range of normatively significant abilities and skills persons develop, and the many (including non-conceptual) ways they engage in the normative domain.

---

[46] Lavin's own conclusion with respect to the imperatival interpretation is that 'it is not forced on us by argument' (Lavin 2004, 428), though he is careful to say that making that claim does not settle its truth.

B. The Many Ways of Going Wrong

An obvious objection can be raised at this point by the advocate of an understanding of the error constraint as a matter of violable / violating principles. The advocate can ask: ok, let us imagine we do not so confine our understanding of the possibility of error – what, then, is error? How do you account for making mistakes, of going wrong, in the absence of such principles? We must confront this question, but before we do so let us make the following observation: the objection is posed in absolute terms, i.e. it is assumed that either error-making is about principles or it is not. In fact, we need not be so absolutist about it: principles (or rules, norms, etc) can continue to play an important role in the possibility of error argument. The point is not to restrict that argument to them, and thus not expect too much from an account of the nature and function of principles. In order not too expect too much, it is not necessary to get rid of principles altogether. The argument – as throughout this thesis – is simply against the necessity of recourse to principles. Principles – like any other normative resource – are not necessary; they are contingently useful, but can sometimes be experienced as necessary. The difficulty, as so often in philosophy, lies in constructing simple and clear examples that can show us the contingency of normative resources, but also to deconstruct those examples that appear to show the necessity of determinants of appropriateness and (in this Chapter) inappropriateness.

Let us, then, try to show, with examples, that we can make sense of the possibility of error in the absence of principles. To help us in this task, let us turn to a paper by Paul Roth entitled 'Mistakes' (2003). Roth does not mince words; early on in the paper he asserts: 'Mistakes can occur even in the *absence* of rules, norms, etc' (Roth 2003, 390; original emphasis). To see the difference between a case where a mistake is said to be made because it is a violation of a rule, and one where this need not be the case, consider the difference between the following two examples given by Roth. In the first, Roth's daughter, Emma, displays her counting abilities to him, at first correctly enumerating the objects before her, but then continuing to increase the count, pointing again and again to the same objects (Roth 2003, 390). This is a case of a mistake, says Roth, that we can understand on the basis of it violating 'a rule or norm that partly constitutes what "counting" is for us' (Roth 2003, 390). Contrast this case with the following: watching the real TV show, 'Blind Date', where contestants, behind a screen, answer each other's questions and try to impress each other. In this

case, says Roth, 'we possess no theory marking out, e.g. which actions must impress and which not, and experience indicates we can be surprised on this point' (Roth 2003, 390). In this second case, Roth is pointing to the possibility (though he does not elaborate on it in this way) of an ability to appreciate that something is not impressive (to a potential date) that does not depend on that appreciation being a claim or assertion that that someone has violated a rule or norm (of how to impress a potential date).

It is important to use the language of appreciation here (as opposed to, for instance, the language of justification or systematisation). This is because situating our descriptive efforts only within a context of justification or systematisation (let us stick with justification, for simplicity's sake) – i.e. where we would be pictured as making a claim or assertion that we would need to back up on some ground – would tend to invoke a need or dependence on rules or norms. By contrast, a context of appreciation does not require the making of a claim or assertion, and thus enables us to loosen the hold of that picture of cognition according to which our evaluation of things is dependent on being guided by rules or norms (that determine how things ought to be). Certainly, a context of justification can be introduced into the scene: for instance, we may be challenged by our partner on the couch as to why we winced when we did, and under that pressure, we might formulate our appreciation as a reason that takes the form of some rule or norm that we think our partner might treat as justifiable, e.g. we might say, Aidan violated the rule or norm for being too cocky, too sure of himself; or, in being too cocky, Aidan violated the rule or norm that if one wants to impress a potential date, one ought to appear at least a little vulnerable and modest (imagine that we know, say, from past experiences of how our partner has spoken of impressing the opposite sex that there is nothing worse than being too cocky).[47] This can happen, but it is does not happen necessarily. That fact that it can happen, but need not, can be difficult to see. It becomes difficult to see when we forget about the contingency and artificiality of a context of justification, and we treat the scene (of evaluation) as one where the persons making the evaluation are implicitly situated in such a context (and thus perpetually subject to making justifiable

---

[47] In fact, this scene is already interesting from a second-person perspective, for here what one of the parties may appeal to as a justification may not be how they saw it, but how they think the other might approve of them seeing it (here we see how a second-person relation can trump, even in a context of justification, what is at stake, i.e. what is at stake is the quality of the relation between the couple on the couch, and not the independent status of a judgement).

claims). Further, it can be difficult to see when the case is described in such a way where it will seem that the wince (at Aidan's cockiness) could not have been made – would not have been possible – in the absence of a previously existing rule or norm, and thus come to depict the wince as explicable on the sole basis that it was either or both guided by and revealed an underlying, previously existing rule or norm (about Aidan doing something, e.g. being cocky, that did not impress the potential date). The problem with describing the case that way is that it makes it impossible to see how one can appreciate Aidan being unimpressive in a way that is not always and already enabled, or revealing of, some pre-existing underlying convention, norm or rule (according to which, on this view, it is unimpressive).

To scramble out of the trap of seeing things this way, one needs to first unravel the description, and return to a very important point made by Roth: the capacity for surprise. We can come to learn new ways of impressing (or not impressing) dates that we did not appreciate previously; after having learnt them, we may very well use them, on occasion, to guide our appreciation of what is impressive or not. In other words, when things are described in such a way that they leave no room for the possibility of being surprised – and for new forms of how not to impress to come into being – then we might start to hear warning bells about how the case is being described. That, then, is the first thing to do: to make room for surprise, and thus for learning. But the next thing to do is to see that it is possible to develop abilities to appreciate that something is impressive or not without this being a case of being sensitive to rules or norms – i.e. that it is possible to learn what might be or might not be impressive without learning rules or norms, and that it is possible to exercise an ability that is not mediated by a process in which we represent some rule or norm to ourselves, consider it relevant to the circumstances, and then apply it to those circumstances. Again, we can better see this possibility when we prioritise the context of appreciation, as opposed to the context of justification (or systematisation).[48]

There is another important warning that needs to be heeded when contrasting the case of counting with the case of impressing a potential date. An advocate of the rule-governed view might say that yes, it is true to some extent that there is room for surprise in the practice of impressing a date, but that is a fortuitous example; take the

---

[48] See Section 26.B. for details.

case of counting: there, surely, you have an example of an activity governed – and necessarily governed – by rules, such that it is not possible to make a mistake in any other way than by violating those rules. We see here the danger of contrasting the two cases (of course, we need the contrast initially to help us see the alternative – the point here is just to avoid hanging on to the contrast). What is important to say, in reply to the advocate of the rule-governed view, is that in order to see that we can make sense of error, and to see that mistakes are possible, without rules or norms, one need not confine such cases to contexts that are (said to be) not rule-governed. Whether rule-governed or not (or more rule-governed / less rule-governed, if one accepts that this can be a matter of degree), it is always possible to find something mistaken that is not necessarily a violation of a rule or norm. Take the case of counting again.[49] For a start, one need not (by necessity) think that Emma continuing to count is a violation of some rule or norm, i.e. it does not follow from an observation that something may be evaluated as a violation of some rule or norm that it is mistaken. For instance, imagine that Emma has recently learnt to count to 10 and that there were 5 objects in front of her – here, we can understand her continuing to count as a way of showing us how far she can count. But second, and more importantly, there are many ways of counting, which could be appreciated as somehow mistaken (including strange, unfamiliar, non-intelligible, and so on) that have nothing to do with any rule or norm already in place, e.g. imagine Emma pauses each time after counting an object and recites the numbers 1 to 10 on her fingers to remind herself what the sequence is, and only then goes on with counting the next object. Here, we would not necessarily say that Emma so doing is a violation of a rule or norm, but we might nevertheless find the behaviour somehow incongruous with the practice of counting.

We still need to dig deeper here, because the advocate of the rule-governed view can always still reply: ok, but Emma so pausing counts as a mistake, or is incongruous, only because it is a violation if not of a rule or norm then perhaps of a convention, or of something more nebulous but still something one can call a determinant of inappropriateness. To stop this kind of thought from overtaking us, we must return to the notion of the possibility of the emergence of novel forms of

---

[49] When speaking of this case, Roth says that Emma violates rules that are 'partly' constitutive of what counts as counting for us, but he does not elaborate on what he means by this 'partly'. This, therefore, is not a criticism of Roth, for the 'partly' suggests that, even in the case of counting, he may be in agreement with acknowledging the limits of the rules, and thereby also with the possibility of appreciating novel forms of inappropriateness.

inappropriateness. The advocate of the rule-governed view attempts to describe these cases as if such novel forms cannot come into existence – that whatever is evaluated as inappropriate, at any point in time, can only be so because of some pre-existing standard or criteria that determines that it is so. But to say this is only, once again, to prioritise the context of justification (or systematisation – again we are sticking here with justification for simplicity's sake). The problem that we face here is one of how cases are described; once described or described in a certain manner, they can make it seem as if what is being used to explain the evaluation that something was mistaken is what pre-existed and guided the evaluation. In a way, the description can turn what was an appreciation into a claim or assertion that is justified / justifiable (and only justifiable) in the form of an appeal to a rule or norm (or some other determinant) having been violated. Trying to avoid this means, in the example of Emma pausing, saying something like this: at no time before have we encountered someone doing such a thing (pausing in between numbers), but this does not mean that it cannot seem to us as quite possibly unhelpful – e.g. possibly unhelpful if we were to adopt it as a guide and standard for counting, for (say) it would not help in the business of counting, but rather make the process slower and more cumbersome (the point is that this is a creative decision, and not something determined, in advance, to be appropriate or inappropriate). Surely, we could come to appreciate pausing as inappropriate, but without this being mediated by some previously existing rule, norm or convention – again, others might disagree with us; our appreciation might not strike them as justified or justifiable; our appreciation is tentative (even to ourselves) – it is not a claim or assertion we are making, and thus holding out to be justified or justifiable. Perhaps, in the course of appreciating Emma pausing as possibly unhelpful we appreciate it in a different way, e.g. we think that it is good that Emma is practicing to count, no matter how she does so, similar to the way that we may not criticise someone for learning to ride a bike by attaching two extra smaller wheels (even if we would not endorse someone doing so for the rest of their biking life; we might hold off, then, on making some claim or assertion about Emma's pausing, and yet appreciate it as possibly appropriate or inappropriate in ways that are novel for us). The tentativeness here is important, but it emerges only when 1) we do not prioritise the context of justification; and 2) we pay careful attention to the description of the case avoiding making it seem like the way we describe is the way things must have / could only have been.

One part of Roth's paper targets the assumption that the normative dimension of social life can only be made sense if we first agree that the object of explanation is one that is necessarily and sufficiently governed by shared rules (or other determinants of appropriateness and inappropriateness). This is an assumption that it will be important to question (we will be doing so in Part IV of this thesis), but it is relevant to notice in this context is that this connects up with Roth's observation as to the many ways in which we can go wrong (ways that do not necessarily include violations of rules, norms etc). Roth articulates these ways partly by reference to Austin's famous paper, 'A Plea for Excuses' (1979 [1956]). It will be instructive to take a brief look at that paper.

In that paper, Austin distinguishes between justifications and excuses. The former are those purported defences of one's conduct – conduct that has been evaluated as 'bad, wrong, inept, unwelcome' (Austin 1979, 176) – that claim that, in fact, one had not done anything bad, etc, but that what one did was 'a good thing, or the right or sensible thing, or a permissible thing to do, either in general or at least in the special circumstances of the occasion' (Austin 1979, 176), as when one justifies murder by saying it was done on a battlefield (of course, knowing what we know from the Chapter above on appropriateness, we know that it is not necessarily the case that if one committed murder on the battlefield that one had done something permissible: such an assumption depends on not contesting the description of the case). The latter defence, i.e. an excuse, accepts that one had done something bad, etc, but one argues that it wasn't the case that one did it, at least not straightforwardly so, as one might say if one was under the influence, or coerced into it (Austin 1979, 176). The interesting point, for our purposes of a discussion on error, is that the distinction drawn by Austin is relevant to our understanding of the different ways in which we can go wrong. In the following passage (a footnote in Austin's text, and the portion of Austin that Roth cites) Austin shows us how the distinction can help us get a handle on mistakes:

> You have a donkey, so have I, and they graze in the same field. The day comes when I conceive a dislike for mine. I go to shoot it, draw a bead on it, fire; the brute falls in its tracks. I inspect the victim, and find to my horror that it is your donkey. I appear on your doorstep with the remains and say – what? 'I say, old sport, I'm awfully sorry, &c., I've shot your donkey by accident'? Or 'by mistake'? Then again, I go to shoot my donkey as before, draw a bead on it, fire – but as I do so, the beasts move, and to my horror yours falls. Again

the scene on the doorstep – what do I say? 'By mistake'? Or 'by accident'? (Austin 1979, fn. 1, 185; quoted in Roth 2003, 392)

Austin's point here (as is clear also from the text to which the footnote is connected) is to point to the ordinary usage of expressions such as 'by accident' and 'by mistake', and show that although they may sometimes appear to be the same, they are not. Austin says that the stories in the above case should illustrate for us what the difference between the locutions is (Austin 1979, 185). He does not elaborate on this, so let us try do so ourselves. It seems that the first story is the one we are more likely to call a mistake: I set out to shoot my donkey, but shot yours instead – but there is nothing (as is the case in the second story) that made me a shoot a different donkey to the one I intended to shoot. In the first story, I shoot the donkey I intend to shoot, but it is, unfortunately, your donkey. In the second story, I do not shoot the donkey I intend to shoot, for the beasts move, and I shoot the beast I did not intend to shoot. Of course, in the second story, it is still the wrong beast, but insofar as we are evaluating my conduct, we will tend to say that in the first story my shooting of the beast is a mistake, and the second it is an accident – though in both cases I shoot the wrong donkey.

Austin's example lets us note a couple of things. First, notice that I can use the term 'wrong' in a way that attributes or purports to attribute responsibility (i.e. blame the person) or I can use it to simply say that an outcome has occurred that is not the outcome that (I think) should have occurred. This is very clear in the above two stories where in both cases I say that the wrong donkey was shot, but where it is only in the first that I am inclined to apportion blame, for I might more plausibly think that in the first case the person who shot the donkey ought to have checked that the donkey he intended to shoot was his and not his neighbour's (the point is that making a case for negligence in the second case would be harder, rather than impossible). Second, notice that the 'should' in what I think should have occurred can also be given different meanings, e.g. it could be said that I shoot the wrong donkey in the sense in which what I do is wrong because it is a violation of the laws that protect private property, or it could be said that I shoot the wrong donkey because it was not the donkey I intended to shoot. Other possibilities are possible too, e.g. it could be said that my shooting the donkey is wrong because it causes the animal pain and I ought not to kill animals (either at all, or these in particular, or not in such a way that

causes them pain). Going wrong here, then, can take a variety of meanings, and this variety is in danger of being missed if we think that we can only ever make errors in the sense that we violate some principles. The die-hard rule-theorist here might say that all of these cases can be described as a matter of violating rules, but this would surely be stretching the use of the term 'rule', e.g. what rule could be said to be violated by me shooting the donkey I did not intend to shoot? And, could I not discover, in shooting the donkey, that this caused the donkey pain, and upon that basis decide to bring into existence (for the first time) a rule against killing animals (or donkeys) in a way that causes them pain?

Interestingly, Roth uses Austin's examples in a different way, and one that allows him to introduce yet another way of going wrong. Roth calls what goes wrong in Austin examples 'ground-level mistakes', and he contrasts them with mistakes that are more whole-scale, and refer to how things cohere (Roth 2003, 392). Roth's distinction here depends on the consequences of going wrong. In the case of ground-level mistakes, these do not require one to change one's 'categories of understanding' – as Roth succinctly puts it, 'alas, the wrong damn donkey was shot' (Roth 2003, 392). In other cases, however, having gong wrong has the consequence that we need to revise how we 'see things as cohering' (Roth 2003, 392). The kinds of examples Roth has in mind include 'a Kuhnian paradigm shift, a loss of religious faith, a disenchantment with an ideology' (Roth 2003, 392). As soon as one decides that one was mistaken in thinking of women fainting as a medical kind, one no longer needs to explain neurasthenia – the entire idea was one that categorised things in a certain way, and called out for something that no longer requires to be done.

One could try to bring Austin's distinction between a mistake and an accident closer to Roth's distinction between ground-level and whole-scale mistakes. One could do this by saying that we can make mistakes in the sense that we do something we did not intend to do, or we can make mistakes in the sense in which our intention to do something is mistaken. We shoot the wrong donkey or we invent neurasthenia (thinking further it needs to be explained), where 'shooting the wrong donkey' means shooting a different donkey from the one we intended (here, our intention was actually correct, in the sense that if we had carried it out, we would have shot our donkey, as we intended, and not the neighbour's). After all, we intended to invent neurasthenia – it is just that our invention was a mistaken one. But the two distinctions – Austin's and Roth's – remain different, for we could still say, with

Austin, that our inventing (or indeed intending to invent) neurasthenia was an accident, e.g. imagine that we intend to invent the medical kind of men fainting, but unbeknownst to us the patients we were examining were women.

In any event, what the above discussion shows is that one can go wrong in more ways than is ever dreamt of by principles – not only in the sense in which the principles are not sufficient to exclude only a certain limited number of ways of going wrong with respect to the principle (i.e. the principles themselves do not dictate even what it means to go wrong with respect to the principle), but also in the sense in which the principles are not necessary to going wrong (we can go wrong, for instance, by doing something we did not intend to do, or appreciate ourselves as having done or be doing something wrong in ways we have not previously evaluated ourselves).

One way in which the rule-theorist could dig in her heels here is to say: ok, but in what sense can we say that, in intending to invent neurasthenia, we went wrong? Went wrong with respect to what? The question is a loaded one because it assumes that the only way of going wrong is a way that can be justified by pre-existing criteria. Imagine, for instance, that it is said (in accordance with this assumption) that medical knowledge can only ever be said to be 'wrong' if it is certified as wrong by some medical authority. For practical purposes, this might usually – and for good reason – be a reliable guide to what ought to count as medical knowledge. But then we might also, and again for good reason, want to leave open at least the possibility of someone – even not a medical doctor – coming across something important that it would be good reason for us to treat as medical knowledge immediately and without the certification of the authority. The point here would be that this possibility would be missed (as a possibility) if we defined wrong as wrong justifiably or wrong legitimately, and thus did not disentangle the concept of error from the more loaded definition of justifiable / justified, and / or legitimate error. The concept of error may be, but is not necessarily, tied to contexts of justification (and legitimacy), and this is because we can appreciate, in a tentative and also necessarily somewhat risky way (for we are not guaranteed that others will agree with us), what is mistaken.

But imagine, now, that the rule-theorist says, ok, so maybe there is a sense in which something can be wrong and yet not be a matter of violating some pre-existing standard, but it still needs to be wrong with respect to something, e.g. perhaps wrong with respect to how things are. If by this the rule-theorist has in mind someone going around randomly saying 'that is wrong', 'that is inappropriate', at anything at all she

comes across, then clearly there is a sense in which the rule theorist is correct. But the very extremity of such a case should make us prick our ears and be careful here. The important point is not to lose sight of the tentativeness of a novel appreciation of things as possibly inappropriate: if we say that this appreciation must be made on some ground (of how things are), we leave open the door for a sub-personal process that is applying (without the person realising it) rules according to which things are as they are (and, in addition, rules according to which those things can be grounds for evaluations of inappropriateness). In other words, by forcing appreciation to be a matter of having a ground we would effectively be re-introducing the context of justification, but just making it tacit or implicit: we would be saying that appreciation is necessarily like a claim or assertion even when it does not realise it is. Of course, we do not want to divorce appreciation from an encounter with the environment: it is a rich experience that can result in the genuine appreciation of things as possibly inappropriate (the trick here is not to make 'genuine' something that we, as observers or describers of the appreciation, would recognise as a legitimate or justified / justifiable appreciation). In a way, the issue is one of description: do we describe the person as appreciating something on a ground that is familiar to us as a ground (for thinking one has gone wrong, or thinking things to be inappropriate)? If so, then one is not allowing room for a novel appreciation of things as possibly inappropriate – a way that creates the ways that we could, in the future, go on to associate as being a ground for something being claimed or asserted to be inappropriate. Again, it helps to think of appreciation as tentative, as relating to the environment in a way in which it is being appreciated how things might be, where appreciating how things might be just can be appreciating them as inappropriate. This reference to tentativeness helps because it helps to avoid structuring appreciation as a claim, i.e. one which first judges things to be a in a certain way and then infers them to be inappropriate (which, again, we are likely to do if we prioritise the context of justification).

This is not the moment for a full elaboration on the context of appreciation, including normative appreciation (this is done in Chapter 26). For the moment, the key is to make room for the possibility of appreciating inappropriateness without this being necessarily dependent on and mediated by some determinant of inappropriateness.

We have come to the conclusion of this Chapter on the possibility of error. We have not scratched the surface of the many ways that we can go wrong, and the

concept of inappropriateness is one that clearly deserves more study than can be dedicated to it here (see also Wide 2009; and for a psychological treatment see Reason 1990). For present purposes, the most important general point is that it is not necessary to associate error necessarily with violating rules. In making this argument, we add it to the one in the previous Chapter where we argued that it is not the case that the very concept of appropriateness requires, by necessity, reference to determinants of it. The next argument to examine, which also attempts to argue for the necessity of determinants, can be referred to as the argument from so-called constitutive rules or norms. Given its importance for the governance view, it is an argument we will be exploring in two of the following Chapters.

Chapter 7. Constitutivism I: The Nature of Games

The third way that the necessity of determinants of appropriateness and inappropriateness can be, on the governance view, articulated and defended is in the form of constitutivist arguments. In order to continue with our criticism of the governance view, we need to tackle this form of argument.

There are, in fact, different varieties of constitutivist arguments, and some are better understood as claims as to the existence of 'super-norms'. Where constitutivism is better understood in this way, it shall be treated in the last Chapter of this Part (Chapter 9; though the issues certainly overlap). In this and the next Chapter, what is of interest is an understanding of constitutivism that proceeds by way of an argument based upon the alleged nature of some activity, which it is argued consists in (or is constituted by) some determinants of appropriateness and inappropriateness (most commonly, rules). A popular illustration referred to in this kind of argument is that of games, and very commonly chess. A popular application of the argument is to language, as when it is argued that language is like a game in the sense that it too is constituted by rules (in this case, the rules of grammar). This Chapter critically examines this argument insofar as it is illustrated in the context of games; the next Chapter looks at the application of the game analogy to language.

As characterised recently by Ariela Tubert (2010), 'the basic idea behind constitutive arguments is that there are certain features constitutive of belief or action' (but here one can also include other domains or activities) 'which guarantee that insofar as we act or believe at all' (or engage in other domains or activities) 'we are committed either to following principles like the law of noncontradiction or the instrumental principle, or to aiming at something like truth or the good' (or whatever may be the rules we say constitute those activities; Tubert 2010, 656). Tubert outlines a number of possible motivations for constitutivism, including: 1) it offers a response to sceptical challenges; 2) it promises 'to establish the normativity of certain rational requirements in a broadly naturalistic way, without appealing to irreducible normative properties external to the agent' (Tubert 2010, 656); and 3) it offers an answer to questions concerning why one should care about certain principles, i.e. it answers the question, 'Why should I care?' (about these principles) by saying 'Because you are already committed to them' (Turner 2010, 656).

Notice that Tubert's list of motivations refers to such attitudes as 'caring' and 'being committed' – in this case, caring about and being committed to constitutive rules. Later, Tubert also refers to 'respecting' the rules, though there is some ambiguity in her view for she sometimes suggests that the object of respect is the status of the rules, e.g. of soccer, 'as being partly constitutive of playing soccer' (Tubert 2010, 657), and sometimes the rules themselves, i.e. respecting the rules of soccer (and not their status). Either way, Tubert's focus on the attitudes (rather than, say, the ontological status of the rules) allows her to be more careful in her description of the importance of the rules: she speaks of the rules, as we saw in the above quote, as being '*partly* constitutive of playing soccer', and she also speaks of respect for the rules of soccer being potentially outweighed by other considerations (Tubert 2010, 657; though here too she only conceives of the possibility of other considerations influencing one to violate, rather than follow, the rules – and thus she does not consider the possibility of the rules being, for example, insufficient to playing or understanding the game). In proceeding to look at other versions of constitutivism, especially in the context of games, it will be useful to keep in the back of our minds Tubert's focus on attitudes because, even if her own explanation of these attitudes is always one that refers to rules, focusing on the need for attitudes may help us make room for looking at games (both participating in and understanding games) in ways that are not necessarily dependent on (their allegedly constitutive) rules.

An important and influential articulation of constitutive rules, which also draws (though with provisos) on the game analogy, is the work of John Haugeland (1998). As most theorists who refer to the constitutivist do, Haugeland traces the theoretical pedigree to the work of John Rawls (1955) and John Searle (1969), though there are, in fact, some differences in the way Rawls and Searle understood constitutive rules ('practice' rules in Rawls's terms). It will be worthwhile to briefly consider (before turning to Haugeland) Rawls's and Searle's appeals to constitutive rules also because they both refer to games.

A. Rawls and Searle

In 'Two Concepts of Rules' (1955) Rawls opposes the summary conception of rules with the practice conception. On the summary view, 'rules are pictured as summaries of past decisions'; they are 'regarded as reports that cases of a certain sort have been found on other grounds to be properly decided in a certain way' (Rawls 1955, 19). On

this view, the very point of having rules 'derives from the fact that similar cases tend to recur and that one can decide cases more quickly if one records past decisions in the form of rules' (Rawls 1955, 22). One needs rules because a principle would be too vague a formulation to serve, pragmatically, as a guide and an aid – there is a need for something more specific, but not so specific as to be inapplicable to novel scenarios. Of course, not only is this formulation difficult to get right, but even if the rule is formulated well, the rule can be misapplied to a certain case (Rawls 1955, 23) – presumably both in the sense that it can be mistakenly thought to apply (i.e. to be relevant) and in the sense that it is applied in a mistaken manner (i.e. what it is presented as justifiable by the rule is not what it can reasonably be said to justify). There is some ambiguity in Rawls's summary conception – indeed, calling it a 'summary' conception already indicates the ambiguity: Rawls presents the rules as both 'generalisations from the decisions of individuals', and thus as a form of description of individuals' behaviour 'as if they were obeying the rules' (Rawls 1955, 24), but also, at once, as guides or aids to be applied in novel circumstances. There is thus a tension between summary rules as descriptions from the third person point of view and rules as guides from the first person point of view. As we shall see in a moment, what Searle contrasts with constitutive rules are rules as guides for conduct (and not summaries or generalisations).

Contrasted with the summary view is the practice conception of rules. On this conception, rules are pictured as 'defining a practice' (Rawls 1955, 24). Here is Rawls's clearest statement as to what a practice is, and how rules (on the practice conception) are related to it:

> It is the mark of a practice that being taught how to engage in it involves being instructed in the rules which define it, and that appeal is made to those rules to correct the behaviour of those engaged in it. Those engaged in a practice recognise the rules as defining it. The rules cannot be taken as simply describing how those engaged in the practice in fact behave: it is not simply that they act as if they were obeying the rules. Thus it is essential to the notion of a practice that the rules are publicly known and understood as definitive; and it is essential also that the rules of a practice can be taught and can be acted upon to yield a coherent practice. (Rawls 1955, 24)

What is instructive in this passage, and what also marks a difference in Rawls's and Searle's treatment of constitutive rules, is that when Rawls speaks of the practice conception of rules, he argues that the rules must be 'publicly known and understood

as definitive; and it is essential also that the rules of a practice can be taught and can be acted upon to yield a coherent practice.' These are quite stringent requirements, and they tend to exclude the possibility of implicit constitutive rules – and certainly of the kind that are, in principle, not linguistically specifiable.[50] There is also a requirement that the persons participating in the practice recognise the practice as defined by those very rules. Again, this suggests that the rules cannot be implicit (in whatever form), for 'recognition' is quite a demanding cognitive condition (it suggests that one has learnt these rules and understands them to have the status they do).[51]

Before comparing this with Searle's account, let us briefly note how Rawls refers to the game analogy to help us understand practice rules. Rawls says: 'the rules of practices are logically prior to particular cases… given any rule which specifies a form of action (a move), a particular action which would be taken as falling under this rule given that there is the practice would not be *described as* that sort of action unless there was the practice' (Rawls 1955, 25). Thus, says Rawls, 'one cannot steal base, strike out, draw a walk, or make an error, or balk' without the rules of the game of baseball; striking out, etc, 'are all actions which can only happen in a game' (Rawls 1955, 25), i.e. in this precise game governed by these precise rules. These rules 'are not guides to help one decide particular cases correctly' (Rawls 1955, 27); they are not 'maxims and rules of thumb' (Rawls 1955, 29); rather, they assert what is logically possible and impossible: 'In the case of actions specified by practices it is logically impossible to perform them outside the stage-setting provided by those practices, for unless there is the practice, and unless the requisite proprieties are fulfilled, whatever one does, whatever movements one makes, will fail to count as a form of action which the practice specifies. What one does will be described in some other way' (Rawls 1955, 25).

It is interesting that Rawls has slipped into talk of description here – this being talk that seems to be better suited to the summary conception. In other words, it is noteworthy that the effect of the games analogy is to say not only that constitutive

---

[50] Rules may be said to be implicit in two forms: first, as linguistically specifiable in principle, and second as not linguistically specifiable. Later Chapters (e.g. Chapter 14) return to the different ways in which rules can be said to be implicit, as well as implicitly followed.

[51] It is important to keep in mind Rawls's own restrictions of the practice conception of rules for, as we shall see later (in Chapter 18), the practice conception has been stretched by other theorists to cover cases where it is not at all clear whether the rules in question are explicit, publicly known, recognised by the participants as constituting an activity, and so on.

rules constitute the game, but that whenever we describe the actions of persons in what we take to be a game of a certain sort, we can only describe the actions they make that are in accord with the rules of the game (notice that as soon as we do this we will miss – because we will dismiss as not relevant – actions that are not describable as being in accord with the rules; this may make us miss the ways games change that has nothing to do with the prior changing of the rules). In any event, before continuing with Searle, let us recall that the rules in question for Rawls, even in the case of games, must be specified rules that are publicly available, taught and known.

Some years later, Searle (1969) offered a similar-looking account, but one which opposed constitutive rules with rules that prescribe actions (and thus did not make reference to summaries or generalisations from individuals' behaviour). On the prescription view of rules, rules determine what is correct or incorrect, whether something is a breach or observance (Searle 1969, 279). Constitutive rules, on the other hand, are rules about counting something as something. These rules, said Searle, are not rules that can be 'kept or broken' (Searle 1969, 281; by contrast, prescriptive rules can be – indeed, must be capable of being kept or broken). Constitutive rules constitute 'an abstract system'; they are not concerned, as prescriptive rules are, with whether 'an abstract system so constituted is correctly or incorrectly realised in some concrete manifestation' (Searle 1969, 281). Searle, like Rawls, also drew on an analogy with games. This is how he did so:

> It is possible to make this distinction in the case of games like chess. There are rules in chess which quite clearly prohibit certain kinds of conduct and require others – rules about the time allowed for a move, for example. There are other rules which quite clearly do not, by themselves, do anything of the kind. The rules of chess include rules about what constitutes castling, check-mate, and so on. In a very large class of games there is a rule or rules defining what it is to win the game. Yet such a rule does not by itself tell us what to do. It does not forbid us to go on moving the pieces about after the game is won, though it may deprive such an activity of its point. (Searle 1969, 279-80)

Notice that there are differences here between Rawls's use of summary rules and Searle's notion of prescriptive rules. For Rawls, summary rules are at once summaries or generalisations from individuals' behaviour as well as guides or aids. Further, they are guides or aids in the sense that they are like maxims or rules of thumb – they

assist persons to apply some general principle[52] where it would otherwise be difficult for them to do so (to work out from scratch, as it were, what some general principle requires). For Searle, as is clear from his games analogy, prescriptive rules 'determine', as he himself puts it, what is correct or incorrect, and these are rules that can be observed or broken (it would be strange to speak of breaking or observing rules of thumb or generalisations from behaviour). Nevertheless, despite this difference in what Rawls and Searle contrast with constitutive / practice rules, that concept (i.e. of constitutive / practice rules) in itself remains very similar: it is a question of logical possibility, i.e. without such rules (e.g. the practice / constitutive rules of chess), according to both Rawls and Searle, there would be no way in which one could play chess or describe persons as playing chess.

As we shall see later, though, when we look at Searle in some more detail (Chapter 19) despite this commonality, Searle does not share Rawls's commitment to the practice rules (Searle's constitutive rules) being publicly known and taught, for Searle wants to allow for the constitutive rules to be implicit, such that persons can causally develop abilities (that form the background) that enable them to act in accordance with the rules (these being rules that constitute the rules of, say, baseball). This use of constitutive rules is problematic, i.e. a large part of the problem is precisely this move to regularise the constitutive rules – to make it seem as if the constitutive rules are descriptions of behaviour applicable by default, as if behaviour itself had to be by default guided by those constitutive rules. This transformation of logical possibility into an order of regularity applicable by default, makes the normative too fixed and rigid and not tentative and flexible enough, leading also to a very passive description of the mind learning and working (because the mind, on the back of this view, as we shall see in detail in Part III, becomes but a receptacle for learning, repeating and reinforcing those rules).

These are issues, including with regard to Searle, which we will deal with later. For the moment, having considered some of the differences between Searle and Rawls, let us return to Haugeland.

---

[52] The 'principle' Rawls has in mind is (a certain general formulation of) utilitarianism – it being the plausibility of utilitarianism that Rawls was concerned with in his 1955 paper.

B. What Haugeland Demands

Haugeland's starting point is to observe that there are two kinds of rule: first, the rule that is 'merely exhibited in what happens', and second, the rule that we understand as 'governing or determining what happens' (Haugeland 1998, 305). Put differently, the first kinds of rules are 'factual: they are just how things turn out or are likely to be'; the second kind, on the other hand, are 'normative: they are how things ought to be, in some relevant sense of "ought"' (Haugeland 1998, 306). If there are to be governing (normative) rules, Haugeland continues, then there must be room for 'discrepancy between the rule and what actually happens, that reflects not on the rule but on what happens, the rule-following – and hence (in these cases) on the rule-follower' (Haugeland 1998, 307). This is all familiar from the Chapters above (including the reference to the possibility of error imagined as a matter of violating the rule), though it is perhaps starker in its association with rules – it does not use a more neutral term, like the ordered or the patterned, but immediately situates the normative as a certain kind of rule, as distinct from another, non-normative, kind of rule.

The distinction between constitutive and regulative (or in Haugeland language, 'governing') rules is a distinction between two kinds of normative rules. As Haugeland summarises it, the distinction refers 'on the one hand, [to] rules that govern phenomena that already make sense as what they are, independently of those rules, and, on the other hand, rules that define or constitute the phenomena that they govern – phenomena which, therefore, could not make sense independently of those rules' (Haugeland 1998, 318).

The best way to get a handle on what job Haugeland wants constitutive rules to do is to see what kind of normativity is going to be insufficient. The first thing that Haugeland says we need to avoid is a regress which would ensue if 'in order to comply with any [governing] rule…it is always necessary first to comply with some other rule of that sort' (Haugeland 1998, 307). There needs, then, to be a more 'basic' kind of compliance with governing rules (Haugeland 1998, 308). But this basic kind of compliance cannot be, for instance, 'biologically evolved normativity' (Haugeland 1998, 308-310) or 'socially instituted normativity' (Haugeland 1998, 310-313), for the problem with both is, essentially, that the conduct as it is performed is the standard, which just means (for Haugeland) that there is no longer any distinction between the merely regular and the normative.

Haugeland's critique here is important, and in this respect his view echoes that of this thesis's arguments against regularising the normative, but Haugeland's solution is one that goes in an opposite direction to the direction advocated by this thesis, i.e. he attempts to make the normative fixed by external, and 'objective' standards. In other words, Haugeland thinks that the normative, in order not to be regularised and thus subsumed under the regular, must be fixed in some other way, but in arguing this he misses out on the possibility of the normative as both non-regularisable and tentative; as we shall see, this does not mean that we cannot treat something (some external standards) as fixed for certain purposes – we just ought not to see the normative as necessarily exhausted by standards said to be necessarily and always treated in that way. The very advantage of de-regularising the normative is to make it tentative and not fixed.

Let us return to Haugeland's arguments against those accounts of the normative that are insufficient. Take the context of 'socially instituted normativity' (ultimately, Haugeland wants to argue that 'constitution' is something 'essentially more than mere institution': Haugeland 1998, 318, so clearly we need to understand more of what he means by 'institution'). Consider, for example, greeting each other: 'what it is to greet someone,' says Haugeland, 'and what it is to be a circumstance in which a greeting is appropriate, are nothing other than what the community members accept and deem as such – which is to say, they are themselves instituted along with the normative practices in which they occur, and by the same socialising process' (Haugeland 1998, 311-2). Haugeland complains about social norms so understood that they destroy 'the distinction between proper performance and getting things right' (Haugeland 1998, 313). This is because the 'getting things right' is internal to the performance, and not on something external to it (Haugeland 1998, 313).

Our focus here is not on Haugeland's reading of 'social norms', but on how he sets up the dialectic so as to make room for a certain characterisation of 'institution' (and, at the same time, the constitutive). Haugeland contrasts the insufficiency of 'institution' in the above sense with what he says 'objectivity demands':

> What objectivity demands, by contrast, is that the 'objects' of objective
> tellings should have a *determinacy* and normative standing independent of the
> performance norms for those tellings, such that the tellings could be
> performed properly, and still get their objects wrong. In other words, there
> must be two distinct types of norm for objective tellings: propriety of the

performances as such, and correctness of the results vis-à-vis their objects. And the objects themselves must stand as independent criteria for that correctness. (Haugeland 1998, 314; emphasis added)

Notice the language of 'determinacy', and the very resolute emphasis on there being permanently 'independent criteria' (not criteria that we can treat as independent for a certain purpose, but criteria that are necessarily independent). Notice, also, the distinction between ground-level and some higher-order level (a distinction that can remind us of that between domain-correctness and 'correctness itself', or internal and external correctness, which we encountered in Chapter 5). The ground-level is already normative – it is not merely regular, for, as in the case of socially instituted normativity, it allows for guidance and regulation – but it is not sufficiently independent of regularity, and thus its normativity is not sufficiently and permanently external, in order to secure 'what objectivity demands' (and, for Haugeland, ultimately also 'the possibility of truth-telling': Haugeland 1998, 308). To help us get a better grasp on what Haugeland is trying to achieve, let us consider the following excerpt – which, though long, will offer us an opportunity to analyse a detailed example:

In a certain community, everyone learns to grock when it is proper – that is, in grock-worthy circumstances – and not otherwise. In so learning, they learn how to tell which circumstances are grock-worthy, how to tell someone is grocking, and how to grock by themselves. And through the mechanisms of conformist normalising, or something like them, they all acquire the normal dispositions of proper grocking. In this way, grocking and grock-worthiness, and the norms connecting them are all socially instituted together. In terms of these institutions, we can understand what it would be for a member of this community to tell grock-worthy circumstances wrongly (or even for everyone to be wrong once in a while).

But could they ever all be wrong in some systematic way? Or, more poignantly, could they all be wrong but one? Until a few weeks ago, say, everyone concurred in telling grock-worthy circumstances. But then poor Pat began to diverge: sometimes everyone grocked but Pat; sometimes Pat grocked and no one else did. Pat protested and persevered against the tide for a while; but, conformism being what it is, everyone agreed again soon enough. Now, is there any chance that, during the disturbance, Pat was right and the others wrong? Would it so much as make sense to imagine that initially they had it right, but then, somehow, Pat led them all astray? Or, could it be that, initially, everyone had it wrong, until Pat finally got it right, and then brought the others around? On what basis could these alternatives even be differentiated, let alone be decided among? It depends.

Suppose, for the first case, that grocking is a special maneuver that is called for, according to the villagers, in a certain circumstance that occasionally arises in a rather free-form folk dance. Outsiders have a hard time telling when these circumstances arise, not to mention whether the relevant dancer grocks; and the villagers are not very good at explaining either one. But – until the incident with Pat, anyway – they never had any trouble telling what was going on, and agreeing among themselves. In this case, it seems to me that mere conformist norms are the only norms intelligibly in play. If all the other villagers consistently disagree with Pat, then they are right and Pat is wrong, in the only sense of right and wrong that I can understand as pertinent. This remains so regardless of whether there is any independent way for *us* to determine which dancing is the "same" as which, or who changed when. That is, it doesn't matter whether Pat changed, the (other) villagers changed, neither, or both, or even whether this makes sense to ask; at all events, the villagers are authoritative about their dance.

Suppose, for the second case, that grocking is pronouncing the word 'yellow', and that grock-worthy circumstances are those which, in the face of something yellow, call for a color-identifying response. If there is anything to objective truth-telling, a case like this had better be different from the folk dance. On the one hand, of course, which sign is to be used for yellow – *'jaune'*, *'gelb'*, a certain handwave – is entirely a matter of public consensus, just like the dance step. There is no way everyone could be wrong about that, or Pat right and everyone else wrong. On the other hand, however, that some particular object *is in fact* yellow is not a matter of consensus: it's an objective state of affairs, quite apart from any and every response to it. Accordingly, once the customs have established that 'yellow' (or whatever) is the proper way to identify the color of yellow things, Pat's systematic responses to that object could perfectly well be right, and everyone else's wrong; or they could all be wrong.

The trouble is, instituted norms alone cannot sustain this distinction. If what it is to be yellow were itself instituted along with the norm for responding properly to it, then the consensus could no more be wrong about what is yellow than about the proper (customary) response thereto. Imagine a charismatic florist who, without affecting any other color responses, convinced everyone that a certain jonquil is sacred, and that the proper color response to it is not 'yellow' but 'scarlet'. In thus determining the general disposition, the charismatic florist would *ipso facto* just (further) determine what it is to *be* "yellow" (and "scarlet"), for purposes of proper color response in this community – and, if the norms are merely instituted, what other purposes could be relevant? By the same token, however, such responses would not be *objective* color reports (that is, not really *color* reports at all). (Haugeland 1998, 315-7)

Let us focus, first, on the first case, that of grocking as a dance. Does not Haugeland's description over-regularise, and thus present as overly deterministic, the community's grocking (as dancing)? Haugeland's language here is already somewhat vague, for he refers to 'conformist norms' (as distinct, for instance, from 'governing rules'), and it is not clear whether conformist norms are just regularities, or more than regularities

but not independent enough from regularity. Whichever reading one gives it, the fact remains that the description is a very deterministic one. To see that this is so, notice that one can describe the case as one where it is not assumed that the community's reaction to Pat was bound to be one – could not but happen to be one – where Pat would be pressured into conforming back to the community norm. To describe the case as if this was inevitable is equivalent to imagining that the norm that the community comes to appeal to as a result of how it reacts to Pat was always and already there. But it is possible for the community to react by saying that Pat's behaviour shows how the dance was always meant to be danced, as if they had been waiting for someone to come and show them how it could be done precisely in that way. This is, of course, also a kind of deterministic answer, for it describes the community's reaction as but revealing another underlying reality (in this way it offers support to the reaction – it is a way of confirming, of treating as non-contestable, the community's reaction). But the very fact that the community could have reacted differently – and thus the very fact that we could describe the case in a way that leaves open the possibility of how the community reacts – should help us see the underlying indeterminacy of the dance. It is this underlying indeterminacy that we want to present as the mark of the normative. In Haugeland's treatment, by contrast, his over-regularisation, or deterministic description, of 'socially instituted normativity', allows him to introduce that which is more 'demanding' in the second case.

Haugeland is opposing, in effect, two kinds of determinacies: the determinacy of 'conformist norms' and the determinacy of constitutive, higher-order, truth-enabling rules (both are, as it were, fixed, though it is only in the second that we can speak of objective correctness and incorrectness). In fact, as we shall see in a moment, matters are more complicated, for although Haugeland wants there to be this higher-order realm of constitutive rules, he does not want these to be hovering in some no man's land – instead, he wants them to be there implicitly, as if giving implicit status to our regular compliance with 'governing rules'. It is as if the implicit existence of constitutive rules gives those practices that we conform to the opportunity to be characterised as 'governing rules' rather than 'conformist norms' (and other regularities). To put it differently: although Haugeland allows for there being some practices that are just a matter of conformist norms (where there is no possibility of truth-telling), he also wants it to be the case that we comply (that is the regularity bit)

with governing rules (these being rules that are now called 'governing' and no longer mere 'conformist norms') the status of which (and here is the strongly normative bit) is guaranteed to be correct by some higher-order (constitutive) level. To unravel this, we need first, as we have done above, to question the description of the first case (which made it look like it was inevitable that the community would react by forcing Pat back into line), and we also need to question the description of the second case, which we are yet to do.

Take, then, the second case. Here, Haugeland wants to say that there is an objective matter of fact about what yellow is, and that it is this standard that determines whether someone uttering the term 'yellow' is correct or not. Let us look at this case a little bit in more detail. What Haugeland's description of the case makes it difficult to see is how there could arise reasonable disagreements. Imagine, however, that we disagree about whether such-and-such an object is yellow. You say it is yellow, and I say it is not. I say it does not appear yellow to me, and you say it appears yellow to you. To try to resolve the dispute, we could do a number of different things: e.g., we could see what a third party (the more the merrier) says about this object, does it appear yellow to them or not? Or, if we did not have the time to do this, we could look for a scientific explanation of yellow and test the object under those criteria. And, if most people said it was yellow and / or the test came back as positive (i.e. the object was yellow according to the criteria), then we might agree that there is something wrong with me, and that objectively the object is yellow and serves as an objective standard. This might be what Haugeland has in mind. But now what is going on these attempts at a resolution? In both cases, are we not referring to the use of the term yellow? In the case of the third party, or a crowd of third parties, we are trying to resolve the disagreement by the criteria of how most people use the word yellow, for it is to this object that they would apply it to. In the second case, we are obviously using a standard – some scientific criteria that we have agreed will serve as the criterion for calling this object yellow. After all, how else could it have happened that we call this yellow, rather than say zello? So the first answer to Haugeland's second case is to say that our disagreement is still about the use of the word yellow: does it apply to this object or not? Haugeland makes it look as if the disagreement is about whether this object really is yellow, and there is a fact of the matter whether the object is yellow. But, of course, yellow, with or without quotation marks, is still a term, and one that members of a community use in various ways –

sometimes to refer to certain objects or describe in certain ways. There is simply no other way to understand how we could try to resolve a disagreement about whether something is yellow or not. But the second answer is that neither the crowd nor the criteria will, by necessity, decide (as if by itself) whether we ought to call this object yellow. To see this, imagine that I tell you that this object appears almost yellow to me, but not quite – it looks zello to me. We have used the scientific criteria, and we have used the crowd of people, and they all say it is yellow, but I still stamp my foot and say this object is almost like yellow, but not quite – it is more like zello. Now the point here is that this case is exactly like the first of Haugeland's cases, but not under Haugeland's description of the ground-level. There are a number of things that could happen. First, it could be said, by the community, that although this may appear differently to me, I need to learn to use the word yellow to refer to it, because that is how we all refer to it and have decided to refer to it. The second thing the community could say is that it agrees to make a new term: we all agree to call how things appear when someone has a doubt about whether they are quite yellow (when some – but not necessarily all – think they look almost like yellow) 'zello'. The key point to see is that there is no determinacy: what the community decides to be relevant or important (in this case, whether it is relevant and important to distinguish between yellow and zello) is not pre-determined – it could go either way, e.g. how useful it seems, or might even by accident come to seem, to distinguish between yellow and zello. At this point Haugeland might say, ok – but surely now we have two terms that together determine what is yellow and what is zello. Any future disagreements will be decided by an objective fact of the matter as to whether something really is yellow or zello. But we would reply: true, but only until we have a new disagreement, and then again we need to decide how to deal with it – of course, we could dismiss it (and just tell the person that they are being fanciful, or that there is something wrong with their eyes etc), but, we may also not, and in not dismissing it we could create a new term.

The vital point is not to make us necessarily passive vis-à-vis that which can emerge as normatively relevant, as if whatever we could ever consider to be correct (or important and relevant) is pre-determined by what really is correct (or important and relevant). Haugeland introduces a division between 'conformist norms' and 'governing norms governed by constitutive rules' because he is afraid of 'anything goes'; of there being a kind of mayhem where everyone could be a Pat, or where only a chaos of disagreements would ensue concerning what counts as yellow. But this

kind of fear pushes him into too stark, too deterministic, a description both of dancing and of calling things yellow. Haugeland sets up two kinds of orders: the order of regularities where we just keep repeating the same movements and if someone 'diverges', we pull them back into line; and the order of governing rules governed by constitutive rules where, if someone stamps their foot, we can silence them forever by claiming objective correctness. The problem is the assumption of, or the need for, an order, and thus an underlying determinacy – as opposed to an underlying indeterminacy, which we may sometimes find it useful to treat as if it were determined, i.e. as if it were so-and-so and could not be otherwise (as if the entire catalogue of possible colours was already exhausted and no new relevant and important distinctions could be introduced).

So far, we have been speaking of 'what objectivity demands', and of contrasting conformist norms with governing rules, but we have yet to hear Haugeland's references to constitutive rules. But it has been important to proceed as we have, for it is after setting out 'what objectivity demands' that Haugeland introduces constitutive rules. Although Haugeland does draw on both Rawls's and Searle's uses (and examples) of constitutive rules (or practice rules in Rawls's terminology), he also introduces some uses of his own. This he does in the form of a quartet of distinctions between types of constitutive rules. It will be important to take a look at these. Looking at them will also help us see how Haugeland uses the analogy with games.

The first distinction is between constitutive regulations and constitutive standards. A constitutive regulation refers to the 'rules of the game', which 'regulate the actions of the players within the game, stipulating what they may or may not, must or must not do' (Haugeland 1998, 320).[53] In terms of the 'normative authority' of these rules, it derives 'from an (often tacit) agreement or convention (mutual deontic commitment) among the players, and is enforced by threat of ejection from the game' (Haugeland 1998, 320). Unlike constitutive regulations, which govern only the actions of the players, the constitutive standards 'govern all the phenomena that occur within the game, and determine what they are' (Haugeland 1998, 320). It can be difficult to see these standards because some of them also govern actions. Some of them (the ones easy to see) are not – such as the rule that says that 'a baseball game

---

[53] According to Haugeland, it is these 'constitutive regulations' that Searle and Rawls have in mind by reference to constitutive and practice rules, respectively.

needs bases and a ball', or that 'chess pieces must be amenable to the player's moves, and must remain where and what they are when not being moved' (Haugeland 1998, 321) – but the ones that are hard to see include, for instance, the rule that 'a baseball player can become a base-runner by first being a batter and then hitting a pitch into fair territory, being hit by a pitch, or receiving a base on balls (a walk)' (Haugeland 1998, 320). This is a rule that wears two hats, according to Haugeland: it is both a rule that governs the actions of the players ('a constitutive regulation'), but it is also a specification 'of what can and cannot happen in the game', thereby something that 'effectively determine[s] what it is to be a base-runner' (Haugeland 1998, 320) – in this sense, it is a 'constitutive standard'. Haugeland argues that not only is it easy to miss these standards, it is also easy to underestimate them – to leave them in the shadows of the constitutive regulations. In showing how he thinks they can be underestimated, Haugeland focuses on the constitutive standards that do not also govern action (e.g. the rules about equipment, etc). To see how important they are, we are invited to imagine versions of chess (called 'semi-automatic' and 'automatic' chess) where the pieces either sometimes move of their own accord, or do so all the time (with only some of the pieces being movable by persons). In all these cases, there would be constituting standards that would be governing how the pieces could move in the game (indeed, what would distinguish them would be precisely these constituting standards) – meanwhile, the constitutive regulations governing the actions of the players could remain the same.[54]

The second distinction is one between constitutive skills on the one hand, and mundane skills on the other hand. A constitutive skill 'is a resilient ability to tell whether the phenomena governed by some constitutive standard are, in fact, in accord with that standard' (Haugeland 1998, 323). In chess, someone who has a constitutive skill is someone who can tell whether any specific move is a legal one. Haugeland puts the matter very strongly: 'Clearly, no one who lacked these skills could play chess: to be a chess player, one must be able to tell not only whether the candidate moves that one is considering would be legal, but also whether the moves that one's opponent actually makes (or attempts) are in fact legal' (Haugeland 1998, 323).

---

[54] There are some puzzles about this distinction between constitutive regulations and constitutive standards – for instance, would not every (constitutive) standard also always have regulative implications, and would not all regulations be describable as standards setting out what must be, with implications for what the players ought to do? – but as they are not central to what Haugeland wants constitutive rules in general to do in his account, we shall leave them aside here.

According to Haugeland, this skill is itself rule-governed, i.e. as he puts it, 'To exercise a constitutive skill is itself to follow a rule' (Haugeland 1998, 323) – so a distinction is here made between making legal moves and evaluating moves as legal, but in both cases the two skills are said to be rule-governed. Contrasted with constitutive skills are mundane skills, which 'are the resilient abilities to recognise, manipulate, and otherwise cope with phenomena within the game, including other players, as required and permitted by the rules – in effect, the ability to engage in play' (Haugeland 1998, 323). The category of mundane skills, as defined by Haugeland, includes a great many things: e.g. the ability to tell what piece is being moved, what squares are attacked by what piece, whose move it is, as well things that characterise a good player, e.g. being able to calculate quite far in advance, etc. The two sets of skills are mutually interdependent: constitutive skills (i.e. the ability to tell whether the moves are legal) depend on mundane skills (e.g. one most know what piece is being moved in order to tell whether the piece is being moved legally), and mundane skills depend on constitutive skills, for 'there could be no telling which pieces are where on the board if there were no such things as chess pieces' (Haugeland 1998, 324). All this is important because Haugeland's strategy is to 'show how mundane skills, understood in this sense as interdependent with constitutive skills, can therefore also be understood as objective – that is, as having independent objects that are criterial for their correct exercise' (Haugeland 1998, 325).

We will come back in a moment to this classification of mundane skills (as interdependent with constitutive skills) as 'objective phenomena' that determine their own correct employment. Let us, first, examine this distinction between mundane and constitutive skills. It is difficult to see why the distinction here between some of the things included under mundane skills and that which is included in constitutive skills is not better articulated in terms of the minimum skills needed to play the game and the skills needed to improve in the game. After all, some of the 'mundane skills', such as being able to tell which piece is what, read like constitutive skills in disguise, for, as Haugeland himself admits, being able to tell whether a piece of wood is a knight is being able to apply the constitutive standard that guides us in making that assessment. Being able to tell whether a piece of wood is a knight seems like a very different skill from, say, being able to calculate far in advance, or any other skill that might help a player be better at chess. Being able to concentrate for longer periods, remembering better, imagining alternatives better, calculating faster etc – all these one might say

are skills that one can improve by playing chess and improve in order to play chess better. It may be that because of this strange grouping of skills that Haugeland finds he has to make the two skills interdependent. Or is there another reason why Haugeland might want to classify some skills that look to be constitutive as mundane? What is it about mundane skills, rather than constitutive ones, that makes them the candidate for Haugeland's notion of 'objective phenomena'?

This question cannot be answered without returning to, but looking deeper into, what Haugeland means by 'what objectivity demands'. The way Haugeland articulates the aim (of satisfying what objectivity demands) is that it consists in understanding 'the objecthood of objects – their standing criteria for objective skills – in terms of their constitutedness' (Haugeland 1998, 326). In seeking to fulfil this aim, Haugeland dismisses a common way of reading constitutive rules, i.e. as devices for 'counting as', which he characterises as counting something as 'something else, over and above what they already were' (Haugeland 1998, 326), as when we introduce rules to count pieces of wood as chess pieces. This is not objective enough for Haugeland, not only because it could not (he asserts) account for the constitution of the objects that it then says can be counted as something else (e.g. it cannot account for the constitution of those pieces of wood), but because it also 'presupposes the objecthood of the objects that are counted as something else, and merely adds onto those objects some new relative features' (Haugeland 1998, 326). Haugeland further develops his complaint as follows (by reference to the example of chess pieces):

> On this line, for example, a rook would just be a certain figurine (or whatever) for as long as the latter is so counted – with all its chess features relative to that counting. But it would really be the figurine itself that stood as the independent criterion for the players' perceptions or actions: they would have to locate and identify the figurine correctly, in order to tell where the rook is, or to move it. In other words, if all constitution were mere counting-as, it would always presuppose, hence could never contribute to, an account of objectivity – which would forfeit the point. (Haugeland 1998, 326-7)

In this passage, Haugeland seems to think that in order for something to be an independent criterion, it must achieve some status of objectivity beyond our (ever) taking it to be one. For Haugeland, the constitutive standard must exist in a way that is independent of whether we take it to be a standard. The need for some independence from an attitude that attributes certain weight, relevance and importance

to a standard is clearly important, but is the matter put too starkly, i.e. as a battle between complete non-independence and complete independence? Could there not be a more tentative, and yet still relatively resilient option in between? In other words, it might be enough – if we allow for degrees of independence – that we could contingently treat a certain standard, when playing chess, as particularly relevant and important, but never so relevant and important that it is so completely independent from our attitudes so as to be non-contestable, un-changeable, non-revisable. A certain standard could be contingently treated as more relevant and important than other standards – but is saying it is completely independent from our taking, and thus autonomous and necessarily followed, not over-stating its contingent relevance and importance?

A passage later in the paper suggests that Haugeland may be better understood as making a phenomenological argument, i.e. as presenting what is phenomenologically objective in this demanding sense – something that, in the course of playing chess, one ordinarily does not question and if one did question would be an obstacle to playing chess. Haugeland says, for example, 'chess players do not first see jumbles or masses of digits (as others might) and then "count them as" chess pieces, any more than you and I first see sense data or coloured surfaces and then "count them as" a tomato. They don't see jumbles or digits at all (and, indeed, they may never have), but only chess pieces' (Haugeland 1998, 329). It certainly seems true to say that there is ordinarily no need for chess players, when playing, to go through some deliberative process wherein the pieces of wood are recognised as chess pieces – indeed, if they did have to do so, then this would make playing chess very burdensome cognitively. But to say that exercising this deliberative, inferential process is not necessary, indeed could not be necessary, every time a chess player plays chess, does not give it a status completely independent from our attitude. For a start, it is possible to have the experience of going through the deliberative process of considering whether such-and-such a piece of wood (or other material) is a chess piece while playing chess. Such an experience is common with collectors' editions of chess boards and pieces, where sometimes the pieces are only distinguished by very minor changes in size, and one has to try to work out which is likely to be the bishop, which the knight, which the rook and so on. Similarly, there are many stories told of persons in prisons building chess boards and pieces out of rice or rubbish – here, where it may be difficult to distinguish the pieces easily (at least at first) one might

indeed (especially if one is the opponent of the person who built the board) keep in front of one's mind which piece of rice is which piece (imagine one does this by keeping track of what piece was on what starting square – certainly, this would take a great deal of mental effort, but the point here is just to see the possibility). We can agree with Haugeland, then, to the extent that he argues that it is an important skill of chess (call it even a mundane skill if you wish, but it reads more like a constitutive skill) to simply see immediately (without mediation by a deliberative process) chess pieces (we cannot even say see pieces as chess pieces, for this would be phenomenologically misleading), and to not come to ordinarily think to question this seeing, but we cannot agree with him if he says: 1) that there will never be occasions on which this deliberative process may need to be exercised; 2) that we cannot, even in the course of playing chess and without much reason, have an experience that we might call 'alienation from practice', as when we experience a knight as a piece of wood; 3) that given that he seems to think we would never question what pieces of wood count as what chess pieces, that he would assert that the size etc of chess pieces can never change over time (imagine, by contrast, that in some community with only one chess set, the rooks all go missing, after which players bring along wine-corks for rooks). There is, in other words, a kind of contingent phenomenological objectivity here that is harmless from our perspective (it is precisely the experience of some normative resource as necessary, which the second-person perspective makes room for), but this is not how Haugeland would want us to read it; instead, he uses the phenomenology to set up some more demanding – necessary, deterministic – sense of objectivity.

Haugeland has still one more card under his sleeve. Having used games so far to illustrate what he is after, Haugeland proceeds to argue that games (on his own view of games) are in fact misleading as an image of what objectivity really demands. Games, he says, 'are made up – they are our own inventions, tailored to our own tastes and purposes'. Surely this is…the antithesis of genuine independence and objectivity' (Haugeland 1998, 329-30). The first thing to note is that this is too strong a statement about games: it seems too strong to associate a game with rules that are 'stipulated, or at least understood, in advance, and that the pieces abide by those rules because they have been constructed, or at least selected, in advance to behave that way' (Haugeland 1998, 330) – this view misunderstands games, and it overestimates the importance of rules (to say that games are only ever imagined, from scratch, by

some group of persons inventing the rules, such that they then only exist in the form of rules being re-enacted or repeated by persons playing the game is implausible, and the next Section will argue why that is so). But, second of all, let us imagine, for a moment, that games (so understood) are not the right image for what objectivity really demands. What does Haugeland have in mind? He illustrates what he has in mind by suggesting we think not so much of games, but of explaining games. In this respect, he asks us to think about what he calls 'empirical chess'. Empirical chess is the pursuit of an explanation of how chess is played, of 'figuring out what it is – that is, how to play it' (Haugeland 1998, 330). Haugeland calls this explanation (when provided) an 'achievement' and a 'discovery' (Haugeland 1998, 331). But in what sense, he asks, could it be an achievement and a discovery? What could dictate whether it succeeded or failed?

Notice that the very questions as posed return us to what we have encountered above: namely, the assumption that there can be something that determines what is appropriate or not (in this case, determines what has succeeded or failed as an explanation; we also have here the stark opposition: success or failure, with no possibility of degrees of acceptability of explanations). What Haugeland finds he has to do in this case – which is one remove from playing chess, for we are now talking about explaining how chess is played – is to presuppose that chess is exhaustively constituted by certain standards and skills, and that thus what keeps the explanation 'honest' (as he puts it, Haugeland 1998, 331) is whether it explains what chess is, or how it is played, by making explicit, by revealing, those underlying standards and skills. In other words, Haugeland in effect posits the nature of chess to be such-and-such, constituted by such-and-such standards and skills, such that it is these that establish what objectivity demands – they create a kind of object against which an explanation can be tested as true or not, correct or not. Games themselves may not be enough, but the image of games as exhausted by rules, as limiting what is and could ever be 'intelligibly possible' in a game, such that 'anything not within those limits is precisely ruled out' (Haugeland 1998, 332), continues to play a very important role, for it is said to be the form in which objectivity exists demanding what it does, i.e. demanding that our explanations of phenomena conform to the ways things are constituted as being. This, however, is not so much an argument for objectivity, as an illustration of an assumption of it – assuming chess to be so constituted, we have (so it is assumed) an object against which an explanation can be objective (or 'honest').

What this fails to see is that by limiting itself to describing persons playing chess by reference to those allegedly constitutive rules, the description could also be 'dishonest', for it could miss precisely those ways in which chess can change where such change is not the result of a prior change in the rules. There just is no way to guarantee the 'honesty' of an explanation, although one can exercise an attitude of great care in one's description of some phenomenon (an attitude precisely that would need to keep in mind the limits of the rules as a guide to description).

Before we conclude this discussion of Haugeland, it is important to add that Haugeland does make some room for change, but in a very narrow way (and indeed in so brief a way that it makes it difficult to assess what he means). Thus, Haugeland says that although he thinks that what is correct and incorrect at the ground level of playing a game is determined by the rules, he wants to leave room for the possibility of revising or repairing, e.g. the constitutive skills (Haugeland 1998, 334). He does so somewhat reluctantly, for he stresses that revision and repair cannot occur casually, for if they are so revised and repaired 'at the first sign of trouble, then nothing is seriously excluding… This is why', he asserts, 'the skills must be resilient: they must be able to stand up to one another, and hold their ground, lest any contentions among them be hollow and inconsequential' (Haugeland 1998, 334). Nevertheless, resilience 'cannot mean complete obstinacy or rigidity' (Haugeland 1998, 334). The 'relevant abilities and dispositions' of skills must be alterable (Haugeland 1998, 335). This is very reasonable, but two things make assessing Haugeland's reply here difficult: first, he offers no example of a change (or a revision or repair of skills) – the closest he comes is to say that 'If you keep seeing rooks moving along diagonals, and bishops along ranks and files, maybe the problem is that you can't tell a rook from a bishop, or have them mixed up; or maybe they aren't rooks and bishops at all' (Haugeland 1998, 335), but this is hardly an example of change – it offers instead a case where someone needs to discipline themselves to conform to the skills (in other words, this is not a case of changing what might count as the skill or skills for playing the game);[55] and second, he holds on to the idea that 'objective skills must be resistant to

---

[55] The point here is that Haugeland holds on to the constitutivist approach to change, i.e. any change in 'the skills' (imagined precisely as constitutive) is a change in 'the game' (Haugeland 1998, 336), but this neglects to see that persons may need to make a creative decision about what skills ought to be seen as part of the game (possibly quite by chance when someone introduces something to playing chess by accident). The whole point is that this 'ought' is not determined in advance; the 'ought' is precisely the defender of – not that which excises – contingent change. Again, no one has the kind of authority of what counts as 'the game' of chess that it is always and already determined what is and

repair, just as they are to revision' (Haugeland 1998, 335), and here it is unclear why suddenly those very same skills Haugeland was speaking of as resilient but not rigid should become immune to 'revision'. Perhaps the best way to read Haugeland is that when those skills are being used as criteria, then they cannot change – i.e. they cannot change while they are performing that function for they must be treated as unchangeable in order to perform the function of criteria (e.g. when we decide to evaluate two explanations of chess on the basis of chess being so-and-so); but this, like Haugeland's argument concerning experienced chess player's skills to see chess pieces immediately and non-inferentially, reads (most plausibly) like a phenomenological argument: it refers to that experience we have when we use something as a criterion for evaluation, and where we thus do not think to question the status (in that moment) of that thing as a criterion. Insofar, then, as Haugeland presents the phenomenological argument in ontological form it is implausible, i.e. he overstates his case when he says that the skills are not only experienced as unchangeable while being used as a criterion but are necessarily unchangeable.

What all of this raises, then, is the extent to which the language of constitution (whether constitutive rules generally, or the mix of constitutive standards and skills identified by Haugeland) is an adequate way to understand games – 'understand' in two senses, both in terms of engaging in games, and for describing the playing of games. In the Section below, it is argued that allegedly constitutive rules are not enough to understand games; we need to re-imagine games and not use them as these caricature images of exhaustively rule-structured activities.

C. Re-Imagining Games

Instructive arguments against the above understanding of games, including chess, have been provided by a number of scholars. To help us critique the constitutivist reading of games, let us take a look at some of them.

The first paper to consider is Hubert Schwyzer's 'Rules and Practices' (1969). As we shall see, we need to be careful about the message of this paper. In his own terms, what Schwyzer shows is that 'what makes chess-playing the kind of thing it is a matter of what sorts of things it makes sense to say with respect to chess' (Schwyzer 1969, 454). We have to be careful because this can be read in a way antithetical to this

---

what is not part of the game (but we can, certainly, treat the game, in certain moments, as being determined by such-and-such rules).

thesis, for it can be said that 'what it makes sense to say about chess' is, as it were, there waiting to be revealed and appealed to whenever anyone does or says something that does not make sense (just that 'what' is revealed is not necessarily a rule – but perhaps something else, e.g. an underlying 'practice'). On the reading being advocated here, it is better to say: what it *might* make sense to say about chess is not pre-determined (whereas the constitutivist treats it as if it is pre-determined). Let us, keeping this in mind, take a look at Schwyzer's instructive and entertaining example:

> Imagine the following. I have now been in Ruritania some time, and can speak the language tolerably well. One day I accompany my host with his family to a certain building, where a large number of people are sitting in a circle on the floor, murmuring among themselves agitatedly. In the center of the circle is a small table with two chairs, and on the table is a chessboard with chessmen arranged as for the beginning of a game. After a while, two men in elaborate clothes enter the room and seat themselves at opposite sides of the table; whereupon those sitting on the floor fall silent and watch intently. The men at the table then proceed, with what appears to be an air of great concentration, to move the chess pieces around on the board according to the rules of chess. It strikes me, however, that they play a rather wild game, and I can see no consistent strategy in the moves of either player. The excitement mounts until, after an hour or so, white mates black. Then everyone present, including both men at the table, shows signs of extreme relief; they mop their brows, smile and congratulate one another. This seems odd, but not especially so; I have seen many odd things in Ruritania.
>
> Anyhow, I am overjoyed that the Ruritarians are familiar with chess; I have brought a set with me from home, and now look forward to many pleasant evenings playing with my host. When we arrive at his house, I show him my chess set and ask if he would like to play. He turns pale, is horrified and appalled, and insists that I put it away immediately. "Blasphemer!" he exclaims. "Did you want to play at chess with me? And did you forge that chess set?" I protest that I meant no harm, and ask him to explain all about chess. After some hesitation he does so.
>
> There is, he says, only one chess set for each community. Chess is enacted once every year by the priests of the community, for purposes of determining the will of the gods. If white mates black, the community and the crops will flourish; if black mates white, there will be trouble. "Does white always win?" I ask incredulously. My host looks puzzled, and again shocked. "Chess is not a duel or a battle," he explains. "It is a sacred rite. There is no winning or losing at all." ("Of course there is," I say to myself: "for white to mate black is for him to win the game!") So, unconvinced by my host, I talk to several of my other Ruritanian friends, and do a little research on my own, consult encyclopaedias, and so forth. I discover that all my host told me is true. Chess is not regarded by any of these people as something that can be won or lost; they do not respond to checkmate as we do at all. Not only does no one in fact congratulate (or admire or envy) the "winner" because he has mated the other, nor console (or pity or despise or ridicule) the "loser" because

he has been mated-no one considers it so much as relevant or appropriate to react in any of these ways. And when I talk to them about these sorts of reactions, they do not understand: "Chess is not a battle," they repeat. I soon discover that the notions of winning and losing have their place only in such things as wars and battles and duels, and that these people do not have competitive games at all, as we understand them. The only way they can construe the proposal "Let's play chess" is to take it as an invitation to play at the sacred rite. It is as if a child in our culture were to suggest that we play (at) Mass, marriage, confession. (Schwyzer 1969, 456-7; internal footnotes omitted)

As noted above, the example is instructive for our purposes because it can be read in a constitutivist friendly way, which also gives us the opportunity to show why that reading is misleading. The constitutivist advocate could say: that's fine, and all very entertaining, but it does not change a thing. For all you have shown is that there can exist a game called chess*; you have simply said nothing about chess – chess (and not chess*) being *our* game. The key in resisting this reading is in the description of the example. Is the example presented as one where our encounter with another way of playing chess reveals the underlying norms that were implicit all along in our way of playing it that we did not realise were there? Or is it presented as one where, as a result of the encounter, we are unsure what playing chess means and need to decide what it might mean (of course, we may also not decide – just leave it unresolved)? On the way the example is described above, the game played by the Ruritarians is already presented as being a different game – the game of chess* and not chess – and this further entrenches the constitutivist reading. This does not mean Schwyzer does not make a valid point: his point is important because it is showing the inadequacy of any one set of rules to dictate what playing the game of chess means. But Schwyzer seems to read this to suggest that there is something that determines what the game is in that culture or community – just that what determines what the game is is not some set of rules, but some other underlying norms (that constitute a practice) that one discovers when confronted by this other way of playing the game. This is important for Schwyzer for it allows him to critique Rawls's reading of practice as one exhausted by rules, for Rawls, to recall an earlier discussion, defines these rules in terms of publicly known and taught rules – and Schwyzer's point is that if rules are understood in this way, then a practice cannot be understood exhaustively in terms of rules. This is not satisfactory for present purposes because it makes it look as if 'what' is involved in playing the game (the 'practice of chess' in Schwyzer's language) is

always and already there, just not necessarily known by us (in the way Rawls requires rules to be known). What we want to say, instead, is that it is only if one describes the case after our taking the game played by the Ruritarians to be a different game (to be chess* rather than chess) that we will think that the encounter reveals underlying norms in our practice (norms not revealed by some set of rules). But the encounter can be described in another way, i.e. precisely in a way that shows that what it *might* make sense (for us) to say about the game (about how the game is to be played) is not pre-determined, for how we react to this encounter with the Ruritarians is not bound to be one (though it can be described / presented as if it was bound to be one) according to which we define that game as a different game (as chess* and not chess). The point is not to let the description of cases present us with a status quo, with forms already made – in this case, chess and chess* – for this presents us as passively confronted with what, in fact, we can actively decide, i.e. are we going to say that the game cannot be played as a sacred rite or not?[56] Is that going to be what we treat as part of the game or not?

The example can mislead us into thinking we only operate with forms already set out for us, neglecting to see that what we treat as a form is contingent, and dependent on our activity and creativity. In this sense, an analogy between the normative and games can be made, but only if we realise that games are indeterminate – that how we might go on to play them is not determined by any pre-existing or underlying norms (not to mention authoritatively laid-down rules). The normative ought to help us see our activity and creativity and not hide it, not define it out of existence, and what we see when we study games is how it is precisely the need for, and the possibility of, this activity and creativity that is at stake in how we depict the nature of games (if we present the normativity of games as tentative and incomplete, then we also see the need for a description of the mind learning and working as active and creative, and not just passively re-enacting actions by participating in forms always and already confronted by us as pre-structured in some ordered / patterned way).

Let us turn next to a paper by Michael Sean Quinn, entitled 'Practice-Defining Rules' (1975). The paper offers some counter-arguments against Rawls's idea (which we encountered above) that there are activities that we can understand as practices,

---

[56] Again it is not the case that we must decide – we could leave it unresolved.

where practices are constituted by, or more generally have some 'necessary connection to rules' (Quinn 1975, 76). Quinn offers six such counter arguments: 1) the argument from cheating; 2) the argument from ineptitude; 3) the argument from strategic considerations; 4) the argument from offenses; 5) the argument from the semantics of rules; and 6) the argument from automatic response. It will be helpful to look at these carefully.

Cheating is an interesting case because it challenges the idea that the only way to engage in a practice (an activity defined as a practice) is to follow the rules. According to the definition of a practice, in order to play the game, one must follow the rules, but cheating (again on this definition) is not following the rules. This suggests that whoever cheats is not playing the game: but would we always want to say that someone who is cheating is not playing the game? A cheat tries to go undetected, so the answer must consider whether it matters whether we evaluate the player as cheating or whether that does not matter. The answer must also take into account whether or not the person knows that he is cheating, for it might be possible that what we evaluate as cheating is what the person thought was a missed opportunity by everyone else. To keep things simple for the moment, let us imagine that the person is cheating deliberately (he is trying to seek an advantage knowing that the way he is doing so is not permitted by the rules), and that we catch him doing so. Imagine for instance that the rules of chess prohibit using caffeine during a game, and what the player does is to use caffeine strips on his arms (which are impossible to detect unless one asks the player to take off his shirt), but we catch him out when, in the heat of the moment, he rolls up his sleeves. We would say the person was cheating, for he was violating a rule, but would we say that this person was not playing the game? You might reply that the rule is not that important a rule, and that the example should involve a 'truly constitutive rule', e.g. moving the knight in accordance with the rule for moving knights in chess. So imagine now that during time-trouble (when I do not have much time left on my clock), I deliberately move my knight two squares forward and two squares to the right, and this gets me an advantage, but not so obvious an advantage that you immediately detect it. We play the game to the end, and I win. Did I play the game, or does my cheating (deliberately violating a rule) mean I did not play the game? This kind of circumstance occurs sometimes in tournament play, and the usual course of events taken is that the result counts if the infraction was not spotted at the time it was made (if it is spotted, then

the person who made the illegal move must play a different move and gets a 2 minute penalty). The point here is just to say that what we might evaluate as the deliberate violation of a rule for some gain (i.e. as cheating) does not automatically mean that we would say that someone – the cheat in this case – was not playing the game. But it also pays to make the case more complicated, i.e. precisely by not describing matters as if the person is evaluated by us to be a cheat, for in the case above of the caffeine strips we might never have previously considered whether it matters whether someone has caffeine in her blood when playing chess: the point is that it is not a foregone conclusion whether this person is a cheat or not.

The above examples are not Quinn's, and indeed Quinn does not provide an illustration. What he does is to distinguish cheating from cases where 1) a person is 'ignorant of some central rule of the game, believes that he knows all the rules, and engages in the practice' (Quinn 1975, 79), as when someone plays basketball without recognising what would normally count as fouls; and 2) a person is pretending to be playing the game (as when someone in a closed room next to ours calls out 'Check', 'Your move' etc to make us think he is playing a game). In the case of the cheater, he knows the rules, he would recognise his behaviour as justifiably criticisable by persons on the basis of the rules, but he does it anyway to get some advantage. Here, Quinn says, rather than saying he is not playing the game at all, we might want to say that the cheat 'is not playing fairly' (Quinn 1975, 80), though he is pretending to do so (recall the requirement of covertness).

The second argument focuses on ineptitude: would we say that someone who is 'so clumsy that he cannot act in accordance with the rules' would not be playing the game (Quinn 1975, 80)? Imagine, says Quinn, a beginner playing basketball: he has been taught the rules, he knows them, he tries to follow them, and yet many familiar with the game would evaluate some of his behaviour as violating the rules, e.g. imagine the beginner bumps into other players (which would normally be counted as a foul). Here, the constitutivist argument draws a very sharp distinction between learning to play and playing: but is it really the rules themselves that determine how we might evaluate when someone is learning and when they are playing? Is it not the case that many of us learn in different ways, and play at various levels of 'confidence' or 'clumsiness', and that this variety is not only not explicable, but also missed, by an exclusive focus of allegedly constitutive rules? Further, might not certain ways in

which someone is clumsy contingently introduce changes to how the game is played?[57]

The third argument (from strategic considerations) suggests that the constitutivist has a hard time considering cases of conflict between the allegedly constitutive rules and strategic rules. Quinn's example is as follows: 'at some points in basketball games, it is strategically wise to foul. This may prevent a basket and reduce the probability of a two-point accumulation' (Quinn 1975, 81). Quinn here adds that 'Knowledgeable experts on basketball agree that such fouling is not cheating' (Quinn 1975, 81). Would we say that players who utilise this strategy are somehow not playing the game of basketball?

Next, as the fourth argument, take the example of actions that are offenses but are nevertheless part of the game (the point being that they ought not to be there if the rules strictly determine what it is possible to do in a game). Take clipping in American football (and imagine that we are not talking about strategic clips or clumsy clips – to distinguish the case from the above two). Here you might say that 'one need not have learned how to clip in order to know how to play football' (Quinn 1975, 82) – but as Quinn notes, this is 'misleading', for in fact in order to play football one must learn to recognise clipping and watch out for it. Perhaps the point here is put more clearly by saying that the occurrence of clipping in a game of football does not automatically mean that football is not being played, even if the rules say that one cannot play football if one clips. Clipping is strictly banned by the rules of football, but the presence of clipping is part of the game – here, again, the rules are not sufficient and not helpful in helping us to either see what persons need to engage in the game and / or how to describe the game.[58]

The fifth argument concerns the semantics of the rules, and Quinn's point is simply that two sets of rules can be extensionally equivalent, even if they express this in different ways; one could then learn to play the game by a different set of rules, and

---

[57] It is important to emphasise that we are not here endorsing the idea (it is not clear whether it is Quinn's) that beginners are but clumsy followers of rules, and experts are those that are confident followers. Rather, we are using his reference to 'clumsiness' to bring out the non-rule structured sense of games. The non-rule structured view of games has important implications for how we understand expertise; the need for re-imagining expertise is discussed later, e.g. in Chapter 13.

[58] Again, we might not necessarily present the case in this way, for presenting it in this way risks us thinking that the game is always such that certain things – e.g. clipping or not clipping – are part of the game, whereas our point is not that we do not, in advance and necessarily, know what is and what is not part of the game. In any event, the example is useful to help us be more modest about claims concerning the explanatory reach of rules.

yet still be counted by onlookers as playing the game. This suggests that one does not have to learn these specific rules – it is not the rules themselves (in their actual garb) that determine what it is to play the game (the point being that the rules always do have an actual garb, i.e. they are limited in how they guide).

The sixth and final argument concerns how much a constitutivist wants to insist on players being guided by the rule. We shall be looking at the problems that arise here in Part III, so let us just mention here that the issue Quinn raises is that if the constitutivist argument requires that the players engage in practical reasoning by reference to the rules (in order for the players to be engaged in playing the game), then this appears implausible, first because it is implausible to think that all actions in a game are the result of practical reasoning, and second because not all 'instinctive actions' can be thought to be the result of some subpersonal process of following the rules (Quinn 1975, 82). Quinn's example here is of a person who has internalised all the rules (however one may want to understand internalised), but who 'has never been guided by one particular rule, because, say, he has never actually run across the relevant situation' (Quinn 1975, 83). The way Quinn uses this example is problematic, for he says that it is possible for the person to act automatically in a way that is in accordance with the rule even though he is not guided by it – and this is problematic, for it neglects to see that no rule applies by itself; there is no sense in which a person acts in accordance with a rule in the absence of someone taking it that applying such-and-such a rule is relevant to understanding (and possibly also evaluating) the behaviour of that person. The more general point simply is that there are many aspects to playing the game that are neither a matter of following or not following the rules, although perhaps some of these could be evaluated by onlookers – if it was thought relevant by them – as in compliance or as in violation of the rules. Overemphasis on the rules makes us see only those abilities that we can straightforwardly describe as rule following or rule violating – and this makes us miss many abilities that may be part of the activity.[59]

Finally, by way of a discussion of the literature on activities – and especially games – that emphasises why the constitutive reading of games falls short, let us briefly note the work of Francis Kew. In a series of papers (1986, 1987, and 1992)

---

[59] Again, care needs to be exercised here, for the point is not that there is a set of non-rule based abilities that determine what it is to play the game, but that no contingent way of treating what is part of the game is exhaustive of the game.

Kew has argued against what he calls a 'formalist' understanding of the nature of games:

> According to formalism, game rules govern game procedures by determining the limits on the ways of achieving the deliberatively contrived ends of games. Players enter into a contract and must submit to authoritative decisions involving the rules, an authority which has an aura of permanency in games such as soccer and rugby union where the rules are termed 'laws'. Hence game-derivative notions such as 'playing', 'winning', 'fouling', 'cheating', 'scoring' get their meaning from the formal rules. (Kew 1992, 294)

According to Kew, the attraction of formalism is that it provides a quick and easy way of understanding some very visible aspects of games. It does so, however, at the cost of representing games as rigid and static, almost 'unreal' as Kew puts it, as if 'the rules provide an ontological break' with everyday life (Kew 1992, 295). It presents games as 'governed by their discursive formulated rules thereby separating them out from other, environing, social processes' (Kew 1992, 295). Kew calls such views 'decontextualised non-sociological analyses' that 'signally fail to capture the complex social dynamics' of games and fail also to account for how and why 'the rules of sporting games are subject to such chronic change' (Kew 1992, 295). Of course, a view being 'sociological' does not guarantee it will capture those dynamics, as Kew points out by referring to research on games inspired by Talcott Parsons.[60]

Kew's critique draws heavily on Harold Garfinkel, and ethnomethodology more generally, and so it focuses on how 'formalised rules', or 'discursively articulated rules' are not only not sufficient for understanding games, but how they construct games as an object of explanation that makes us miss a great deal about what goes on when persons play games. We shall later (in Chapter 18) have occasion to revisit Garfinkel, but let us note briefly here that although Garfinkel's criticism of Parsons – briefly put in Kew's phrasing, a critique based on the inadequacy of understanding social action as 'merely acting in compliance with pre-established formulated rules' (Kew 1992, 299) – is a forceful one, Garfinkel's own answer as to what it misses does not let us get away from the spirit of the constitutivist argument

---

[60] Kew has in mind views such as those by Vaz 1977 and Silva 1981, where the focus is on how to 'modify the reward structure and re-socialise players into formal rule compliance', where this focus is itself dependent on an understanding action as necessarily (and matter-of-factually) either in accord or in violation of the rules (Kew 1992, 297). The point is that it being 'sociological' does not mean it is not at risk of the same deterministic view of games made by reference to rules.

(and, more generally, the governance view) as much as it ought to. This is because, at least at times, Garfinkel employs his method to 'reveal' hidden and tacitly followed norms, and thus although he defeats a kind of explicit-rule-determinism on the surface, he supplements it with a kind of implicit-norm-determinism just below the surface. To the extent, then, that Kew draws on Garfinkel – this is visible throughout Kew's work, e.g. Kew postulates the existence of 'constitutive expectancies', i.e. 'culturally-specific skills and understandings which any agent must possess in order to play particular games' (Kew 1992, 300) – he is of limited use to help us move away from the governance view.

But there are other aspects of Kew's view that are useful to note. Consider, for instance, his examples of what is missed by formalism: e.g. eccentric but permissible actions, such as 'always passing backwards (in soccer); always dribbling towards one's own territory/goal (in hockey); always passing to the same player (in basketball); never kicking the ball (in soccer); always keeping possession of the ball 'too long' (in hockey); never passing the ball…or always kicking the ball (in rugby)' (Kew 1992, 301). Kew colourfully refers to these examples as instances of 'spoilsport' play (Kew 1992, 302), which we can add to the list of Quinn's counter-arguments above (spoilsport behaviour is, for instance, different to cheating behaviour for it is not a violation of the rules, but like cheating in that it is an aspect of the game not explicable on the basis of the rules). Another of Kew's examples is what he refers to as the 'ethos of the game', which includes cases of 'actual rule violations which are implicitly condoned by the "police" of game processes and, moreover, expected and accepted by the players', e.g. 'physical contact in basketball, and the distinction between 'good' (accidental and playful) and 'bad' (violent) physical contacts' (Kew 1992, 302). The point here is that the rules do not allow you to see such behaviour as playful physical contacts between players.

It was earlier mentioned that a large part of the focus of Kew's work is on how games change. One of the ways in which he does this (see e.g. his 1987 paper) is by considering how rules are contested, and thus by analysing the informal negotiations between interests groups, including members of rules sub-committees of governing bodies, referees / umpires, players, trainers, coaches and fans. Of course, the rule-advocate might argue that quite often the procedure for changing the rules is regulated – and in any event, if the only way one changes the game is by changing the rules, then surely this only adds fodder to the rule-theorist's cannon. The answer would be

that no matter how far one tries to regulate some activity (and thus including regulating how the activity's rules can be changed), the manners in which the rules are contested, and how this contest changes the rules, is not exhaustively formalisable: it will always need to refer to forms of interaction between interest groups, as well as abilities not captured by the rules, which themselves change the game – sometimes, but not always (and not necessarily) by changing the rules of the game. But this point, as important as it is, needs to be taken with a grain of salt for, as we have been pointing out, what needs to be defeated here is not formalism (the criticism of which remains important as a way of loosening the rule-based image initially), but rather the idea of an underlying order or pattern of doing things (including playing games), whether this be said to be exhausted by formalised rules or by other set ways of doing things (e.g. implicit norms).

We have now provided a glimpse into the literature on games – in all cases, these being examples of works that have helped us challenge the idea that there is some set of rules (commonly called constitutive rules) that are exhaustive of playing games. We have seen that this literature, though helpful, also needs to be considered carefully, for sometimes its target is much narrower than what is targeted here, i.e. its target is some set of formalised, specified, articulated rules, and not the function to which a rule-based explanation is put (i.e. a function that makes it appear as if what it means to play a game is exhaustible by any specification of how one might treat it as ordered or patterned). But supplementing this literature as we have above, it is hoped that some initial plausibility has been given to the idea that we cannot rely on the constitutivist strategy to understand games, nor allow the constitutivist to draw on games as an analogy as long as the constitutivist understands games in that (constitutivist) way. More generally, it is hoped that this undermining of the constitutivist strategy in the context of games contributes to us moving away from the governance view's insistence that determinants of appropriateness and inappropriateness are necessary.

Chapter 8. Constitutivism II: Language as a Game

It was mentioned at the outset of the preceding Chapter that one of the most popular applications of the games image is to language; put differently, language is often understood on the model of a game, where, of course, a game is understood to be a domain constituted by rules. Although we cannot possibly do this way of understanding language justice in this Chapter – this thesis is not a thesis about language – it is important, especially given its popularity and importance in the literature, that we consider it, even if only briefly. Again, just for clarity's sake: this Chapter does not seek to make a contribution to the philosophy of language or the study of linguistics. But given that the context of language has been just as prominent as the context of games in illustrating the case for constitutivism, it would be neglectful of us not to consider it. The aim, in short, is to analyse the form of the argument being made in this context, and not to theorise about language – though, of course, we need to at least make plausible the possibility that language can be understood in ways other than the constitutivist strategy asks us to understand it.

To give us a quick leg up on the spirit of the view we will be examining in this Chapter, consider the following succinct statement by Von Wright (whose work on norms was discussed in Chapter 2):

> Playing a game is a human activity. It is performed according to standardised patterns, which can be called moves in a game. The rules of the game determine, as I shall say, these moves or patterns – and thereby also the game 'itself' and the activity of playing it. We could say that, when viewed from the point of view of the game itself, the rules determine which are the correct moves, and when viewed from the point of view of the activity of playing, the rules determine which are the permitted moves. It is understood that moves which are not correct are prohibited to players of the game, and that a move which is the only correct move in a certain situation in the game is obligatory when one is playing the game.
>
> The rules of grammar (morphology and syntax) of a natural language are another example of the same main type of norm as the rules of a game. To the moves of a game as patterns correspond the set forms of correct speech. To play or the activity of playing a game corresponds speech or the activity of speaking (and writing) a language. Of a person who does not speak according to the rules of grammar, we say either that he speaks incorrectly or that he does not speak that language. The grounds for saying the one or the other are very much the same as the grounds for saying of a person either that he plays a game incorrectly or does not play it at all. (Von Wright 1963, 6-7)

We see here a very clear statement of a certain understanding of games and an application of this understanding to language. Von Wright is not entirely convinced of the application, as when he says, later in the same passage, that 'the rules of grammar have a much greater flexibility and mutability than the rules of a game' (Von Wright 1963, 7). It is instructive, however, that Von Wright follows this observation with the idea 'that what the rules are at any given moment in the history of a language may not be possible to tell with absolute precision and completeness' (Von Wright 1963, 7), for this suggests that the rules are always there – it is just that we have not yet made them fully explicit. This is a common strategy that accompanies the constitutivist argument, and indeed most arguments for the necessity of determinants of appropriateness and inappropriateness. It is as if, in order to defeat counterexamples that show the known rules are not enough / not relevant, the theorist simply stamps his or her foot and insists that even if we do not know about, the rules were always and already there, awaiting our discovery.[61]

---

[61] It is sometimes said that Wittgenstein presents a rule-governed view of language, and one that treats language as a game, or as many games we play with language. This view is surprising, for Wittgenstein often focuses on rules to point out their insufficiency both for engaging in and for understanding activities. To cite just one example, consider the following from the *Blue and Brown Books* (1958): 'For remember that in general we don't use language according to strict rules – it hasn't been taught us by means of strict rules, either. We, in our discussions on the other hand, constantly compare language with a calculus proceed according to exact rules. This is a very one-sided way of looking at language. In practice we very rarely use language as such a calculus. For not only do we not think of the rules of usage – of definitions, etc – while using language, but when we are asked to give such rules, in most cases we aren't able to do so. We are unable clearly to circumscribe the concepts we use; not because we don't know their real definitions, but because there is no real 'definition' to them. To suppose that there must be would be like supposing that whenever children play with a ball they play a game according to strict rules' (Wittgenstein 1958, 24-5; see also Albritton 1959, where this passage is quoted and discussed (at 846), and where quotations of Wittgenstein on rules as well as criteria more general are provided). But Wittgenstein's critique of rules is sometimes read as replacing the insufficiency of rules (explicit, formal rules) with other determinants (supposedly implicit) of the practice of language, e.g. norms or conventions, or as Wittgenstein sometimes put it, 'uses' or 'customs', or 'forms of life'. It is true enough that Wittgenstein opens himself up to that reading, for he sometimes gives the impression of painting a picture of human life as one of participating in a battery of pre-determined paths for speaking, thinking or acting (precisely those uses, customs or forms of life). But there is another way to read Wittgenstein. This is a reading that argues that Wittgenstein's focus was on our *practices of explaining* how we speak, think and act. On this alternative reading, he shows us that when we explain how we speak, think, or act there are certain things that function in our explanation that we do not think to question, and that, as we rely on these things in this way (not questioning them), they serve as hinges upon which we (as it were) swing (this image of hinges is prominent in *On Certainty* (Wittgenstein 1979). The mistake, then, in reading Wittgenstein can come in two ways: first in thinking that he is pointing to things that we necessarily take for granted when we speak, think or act, such that we are always and necessarily guided by something (and that that something deserves to be thought of as a normative resource, though an as-yet not articulated one); and second, in thinking that those things he is pointing to (that we allegedly always and necessarily take for granted) are somehow timeless – that they are those things human beings (being what creatures they are) always and necessarily need to take for granted. Instead, Wittgenstein is better read as recreating (performing) for us, through his specific philosophical style, the dynamics of contestation and doubt *at the level of explanation* – of precisely the interplay between questioning one thing when taking the

Let us now proceed with care, analysing the constitutivist strategy in the context of language by reference to two highly influential contributions: first, by looking at one of the most challenging and sophisticated papers in this line of argument, namely Wilfred Sellars's 'Some Reflections on Language Games' (1956); and second, by briefly considering (some aspects of) Noam Chomsky's view as to the nature of language.[62] In offering a critical reading, it will be helpful to draw on some of the literature that has expressed doubt about the thesis that language can and ought to be understood on the model of a game (at least where games are themselves understood in a certain way). Towards the end of this Chapter, we will return again to constitutivism in general.

A. A Rule-Governed System of Expressions

That Sellars has been enormously influential in shaping the interests and concepts of contemporary analytical philosophy – especially the literature on normativity, including social normativity – is clearly visible in how his work echoes in important contemporary philosophers, such as Brandom, McDowell and Burge.[63] That his work is only likely to grow in stature is further illustrated by the fact that a recent special issue of the *International Journal of Philosophical Studies* (18(3), 2010), dedicated to normativity and naturalism, focuses largely on Sellars. Sellars is a unique figure, for he manages to speak in ways that those sympathetic to Kant find attractive, but also to offer something of an olive branch to pragmatists (see, e.g. the recent discussion of

other to be unchangeable, and vice versa – and not as establishing (or wanting to establish) that at the level of speaking, acting or thinking we are always and necessarily guided by something we take for granted, or that that thing that we are allegedly necessarily and always guided by is in some sense fundamental or foundational and corresponds to human nature. Put differently, one could argue that Wittgenstein's text(s) perform for us the perils of putting the explanatory cart before the descriptive horse: he reveals to us how our practices of explanation tend to make us describe speaking, thinking or acting as if it were always and necessarily guided by something. Of course, the above would need to be supported by serious and patient exegesis of Wittgenstein's text, and the surrounding literature – a task that falls outside the scope of this thesis (Medina 2010 may assist in developing this reading).

[62] Some of the issues raised by Sellars and Chomsky are better analysed as problems to do with how far models of rule-following ought to inform our understanding of human cognition, which is dealt with in Part III – indeed, Chomsky's view (as elaborated on, though not without changes, by Pinker) is returned to for precisely this reason in that Part in Chapter 11 – but they also raise other problems that are pertinent to our evaluation of the constitutivist strategy, and thus appropriately considered in this Chapter.

[63] Discussing Brandom's work falls outside the scope of this thesis. McDowell is discussed in Chapter 12. Burge is not discussed in this thesis; that he has been influenced by Sellars is clear from his oeuvre, but what is also clear is that he has been careful to distinguish his view from Sellars's– for Burge's most recent statement as to the similarities and differences between his view and that of Sellars, see Burge 2010, 433-435. Of course, by no means are we suggesting that Brandom and McDowell adopt Sellars wholescale; indeed, Chapter 12 discusses how McDowell himself distinguishes himself from Sellars.

Sellars in Bernstein 2010). Tracing, in the requisite detail, Sellars's influence on the contemporary philosophical analysis of normativity is outside the scope of this thesis, but it is important to take note of it. Nevertheless, it is not only because of Sellars's influence that we need to consider his work; it itself is full of important insights that we need to analyse.

Sellars's principal aim in 'Some Reflections on Language Games' (1956) is to explicate and defend the thesis that 'a language is a system of expressions the use of which is subject to certain rules' (Sellars 1956, 204). Most of the paper is structured in a way that involves Sellars working through a number of objections to the above thesis. This allows him to avoid what he considers misunderstandings of the thesis, while also arguing for its plausibility. Following this structure will enable us to get at the details of not only Sellars's view, but also some of the views that tend to accompany the above general definition of a language. The first attempt at an objection to the thesis that Sellars tackles goes as follows:

> *Thesis*. Learning to use a language (L) is learning to obey the rules of L.
>   *But*, a rule which enjoins the doing of an action (A) is a sentence in a language which contains an expression for A.
>   *Hence*, a rule which enjoins the use of a linguistic expression (E) is a sentence in a language which contains an expression for E, – in other words a sentence in a *meta*language.
>   *Consequently*, learning to obey the rules for L presupposes the ability to use the metalanguage (ML) in which the rules for L are formulated.
>   *So that*, learning to use a language (L) presupposes having learned to use ML. And by the same token, having learned to use ML presupposes having learned to use a meta-metalanguage (MML) and so on.
>   *But* this is impossible (a vicious regress).
>   *Therefore*, the thesis is absurd and must be rejected.
> (Sellars 1956, 204; original emphasis)

Making sure that the possibility of a vicious regress does not arise is, according to Sellars, perhaps the greatest threat to his thesis, so Sellars proceeds very slowly in rebuffing it. First, he considers and rejects one obvious way of rebuffing it, namely by substituting in the *Thesis* above the phrase 'learning to obey the rules' with 'learning to conform to the rules' (Sellars 1956, 204). This would yield the view that 'A person who has the habit of doing A in C would…be conforming to the above rule even though the idea that he was to do A in C had never occurred to him, and even though he had no language for referring to either A or C' (Sellars 1956, 204). If we did this,

we would get rid of the need for a metalanguage (ML; and thus the possibility of a vicious regress); one would simply learn to conform to the rules of the language (L). This is a way of rebuffing the objection, but it comes at a great cost (too great a cost for Sellars) – precisely as a result of neglecting the distinction between obeying a rule and conforming to a rule. After all, asks Sellars, could we really understand playing a game as a matter of merely conforming to rules? Surely, says Sellars, there has to be a sense in which the moves one makes count as moves in a game – and if that is so, does that not require something more than mere conformity – some sense in which each move must be 'present to mind' (Sellars 1956, 205)? Answering the objection by appealing to conformity to the rules is a way of dodging (or rather, attempting to dodge) a bullet that cannot be dodged. One must confront the question: how can one give an account of obedience (and not conformity) to the rules of the game, but not fall into a vicious regress?

Sellars argues that there is a second way of rebuffing the objection – also ultimately unsatisfactory (but illuminatingly so) – which consists in saying that when we learn a game we 'become aware of a structure of demands (which may or may not have found expression in a language) and to become able to realise these demands and [be] motivated to do so' (Sellars 1956, 206). The difference here with the 'learning to obey the rules of the game' formulation of the thesis above is that it does not refer to the learning of 'verbal formulae'; instead, it refers to the becoming aware of demands, where these demands happen to be those demanded by the rules of the game. As indicated above, Sellars also thinks this reply unsatisfactory, though it does help him clear the path for his own account. The reason why it is unsatisfactory is that it relies on the concept of 'becoming aware', and it leaves this concept unexplained. In effect, what this move does is to hide the problem of obedience under the umbrella of awareness; it does not confront the question, and so we are (almost) back to where we started.

'Almost' because, in fact, Sellars argues that this way of rebuffing the objection 'points the way to a solution' (Sellars 1956, 207). The attempted rebuff focused our efforts on the need for an account of learning a game in such a way that one learnt 'to do what one does because doing these things is making moves in a game (let us abbreviate this to "because of the moves (of the game)") where doing what one does because of the moves need not involve using language about the moves' (Sellars 1956, 207). The problem with the attempted rebuff was that it

introduced the notion of awareness, which, says Sellars, also re-introduced the potential for a vicious regress. The question then (of how we can be said to obey the rules but not end up in a vicious regress in our explanation) is more accurately posed as follows: 'how could one come to make a series of moves because of the system of moves demanded and permitted by the rules of a game' (Sellars 1956, 207), but without appealing to awareness, and thus without raising the potential for vicious regress?

Here, Sellars takes a crucial step. Recall that a moment ago when saying why the first attempt to answer the objection was unsatisfactory, Sellars helped himself to the distinction between obedience and conformity. Indeed, it was on the basis of that distinction that Sellars said that what we need to understand (if we want to understand how we play games – or indeed speak a language) is obedience rather than conformity. The crucial step that Sellars takes now is to say that we ought not have set up a dichotomy between 'merely conforming to rules' and 'obeying rules': this, he argues now, is a 'false dichotomy' (Sellars 1956, 207). It is a false dichotomy because it makes us think that the only way for one's action to accord with the rules is when it is mediated by some intention to so conform; or, put differently, it makes us read 'obedience' in too demanding a fashion, requiring awareness or intention to obey. The second attempt to rebuff the objection fell into precisely this trap, for it tried to base the answer to the objection on some notion of awareness. Thus, what we have to do is find a way of articulating how we can learn to (and actually make) moves in a game, where our learning and doing so is not mediated by awareness or, more broadly, an intention to act in accordance with the rules.

It is not easy to keep track of Sellars here because he has at least two balls in the air: the first one is to keep the rules in place (i.e. of language as a system of expressions the use of which is governed by rules), but the second one is to keep those rules at a distance by giving an account of how persons can act within the system (and thus make moves in the game) while nevertheless not falling into a vicious regress (which ensues the moment one requires some intention for one's acts to be in accordance with the rules) and without imposing a strict dichotomy between mere conformity and obedience. Can such a feat be achieved? Let us continue with Sellars to try to unpack his idea.

In trying to make room for a concept that can pinpoint the sense in which one can make moves in a game but nevertheless not intend to do so (and yet not be mere

conformity), Sellars draws on the image of bees dancing. Imagine a bee returning from a clover field: it twists and wiggles in a certain way (or so says Sellars). These 'turnings and wigglings occur because they are part of a complex dance', and yet, we would not want to say that 'the bee envisages the dance and acts as it does by virtue of intending to realise the dance' (Sellars 1956, 208). We can agree with Sellars that the bee does not envisage the dance, but in what sense then do the turnings etc occur because they are part of a complex dance? The 'because' here implies that there is a gap between one thing and another thing that is joined up in some way: Sellars seems to want us to imagine this gap for a moment, and then immediately let it go – to say, perhaps, the way this thing is joined up with another is that it is a sign of it. After all, the only other way to read the example is to say that we, the observers, are understanding the bee to be making a certain kind of dance, and that is why we understand the turnings and wiggles as moves in a dance – there seems no longer to be any sense in which the bees are connected to this particular pattern / this particular dance (connected in some internal way). But to say this would be collapsing the 'because' into mere conformity, which is not what Sellars wants. What Sellars seems to want is for us to understand this 'pattern' or this 'dance' as severable from any particular representation of it that may figure in our explanation of the bees' behaviour, i.e. to think of it as existing of its own accord, such that the bees' turning and wiggling is, in some sense, an objective matter of being in accord with that pattern or dance. But then if this is what he means why is it that what we are talking about here are not (for instance) laws of science or mere generalisations – in what sense are we still talking about rules?

The viability of Sellars's proposal rests on setting up a distinction not between mere conformity and obedience, but between what he calls pattern-governed behaviour and rule obeying behaviour.[64] Recall that mere conformity to rules was defined as 'doing A in C, A* in C* etc, where these doings "just happen" to contribute to the realisation of a complex pattern', and obeying rules was defined as

---

[64] As shall become clear below, this is a distinction but not an opposition, for the idea is that rule obeying includes, but is not reducible to pattern-governed behaviour. Our disagreement with this, in brief, is that it still attempts to situate the normative within a patterned context. Rule-obeying behaviour is itself, as a result, subject to being described as regularised (the attitude involve in obeying rules is said to be regularly exercised, and somehow attached to underlying regularly performed behaviour). The argument here is that this twin regularisation of the normative (or, more accurately, of rules) gives us misleading pictures of the normative, which also has deleterious implications for our understanding of the social (including our abilities to participate in it).

'doing A in C, A* in C* etc, with the intention of fulfilling the demands of an envisaged system of rules' (Sellars 1956, 207). One of the problems with the distinction was that it offered itself as a dichotomy, so that we could not see how acting in accordance with the rules was possible without intentions to fulfil them. This new distinction avoids this by having one thing involve the other, i.e. by having rule obeying behaviour involve pattern governed behaviour and yet not being identical to it (Sellars 1956, 209). So how does Sellars account for this 'involving yet not identical' relation? Here is one instructive paragraph, which also extends the account to language:

> To learn pattern governed behaviour is to become conditioned to arrange perceptible elements into patterns and to form these, in turn, into more complex patterns and sequences of patterns. Presumably, such learning is capable of explanation in S-R[stimulus-response]-reinforcement terms, the organism coming to respond to patterns as wholes through being (among other things) rewarded when it completes gappy instances of these patterns. Pattern governed behaviour of the kind we should call 'linguistic' involves 'positions' and 'moves' of the sort that would be specified by 'formation' and 'transformation' rules in its meta-game if it were rule obeying behaviour. Thus, learning to 'infer', where this is purely a pattern governed phenomenon, would be a matter of learning to respond to a pattern of one kind by forming another pattern related to it in one of the characteristic ways specified (at the level of the rule obeying use of language) by a 'transformation rule' – that is, a formally stated rule of inference. (Sellars 1956, 209)

This helps us a little bit with the notion of 'pattern governed' – though we have to prepare ourselves to tackle the explanatory value of a hypothetical explanation in terms of a new category of rules ('formation and transformation rules') – but where does it leave obeying behaviour? Obeying behaviour is said to involve pattern governed behaviour, but not be identical with it: so what is the 'more' that it has that pattern governed behaviour does not? It cannot involve awareness or intention to realise or fulfil the rules – for that would be to re-introduce the vicious regress as well as the dichotomy between mere conformity and rule obedience. If not that, then what?

At one point, approximately half way through the paper, Sellars returns to the notion of rule obeying behaviour. He says of it that it 'involves a distinction between a game and a metagame, the former, or "object game" being played according to certain rules which themselves are positions in the metagame' (Sellars 1956, 214). Furthermore, Sellars adds, 'we have emphasised that in an object game played as rule

obeying behaviour, not only do the moves exemplify positions specified by the rules (for this is also true of mere pattern governed behaviour…) but also the rules themselves are engaged in the genesis of the moves. The moves occur (in part, and in a sense demanding analysis) because of the rules' (Sellars 1956, 214). What might it mean for the 'rules themselves to be engaged in the genesis of the moves'? To ask this is, in effect, to ask precisely what rule obeying behaviour is. In effect, we have not progressed any further.

Sellars makes three additional points in trying to give some sense of what rule obeying behaviour might be. The first, and most important, point requires us to cite the following example:

> I am looking at a chessboard set up in a certain way. This acts as a stimulus for the language entry transition into the rule language position '…and my king is checked by his bishop'. I then make the move in the rule language via the auxiliary position 'If one's king is checked by a bishop interpose a pawn' or 'one should interpose a pawn' to 'Sellars, interpose a pawn!' (or correspondingly on the alternative formulations of the auxiliary sentence). The latter is a motivating position in the rule language, and I make the language departure transition from the rule language to the action (in the chess game of interposing a pawn). (Sellars 1956, 215)

This passage introduces some technical terms, but we can get at the heart of the point without them. There are, Sellars says, in further elaborating on the above example, two different things: first, there are such expressions as 'bishop', 'My bishop checking his king', and 'interpose a pawn!', and there are such things as 'piece of wood of such and such shape', 'There is an open diagonal space between this white piece of wood and that red piece of wood' (here we learn the somewhat idiosyncratic colours of Sellars's chess set!), and 'Place this piece of wood between these two!', respectively (Sellars 1956, 215). The heart of the point, then, appears to be that rule obeying behaviour does not consist in first seeing a piece of wood of such and such shape and over there another piece of wood of a different shape, and then responding to it by seeing it as a bishop checking a king – to say this would 'make the word "bishop" [and 'king', and the relation of 'checking'] a metalinguistic word' (Sellars 1956, 216). Instead, the experience '"this is a bishop checking a king" must be a response to a chessboard arrangement, and not to words describing the arrangement' (Sellars 1956, 216). This point is a valuable one in that it can remind us of Haugeland's phenomenological observation: chess players can see chess

arrangements, and not some physical arrangement they infer to be a chess one. In that respect, it is true that I do not, in seeing that the bishop is checking the king, also make a move in the metalanguage: I am focused on responding to a chessboard arrangement (and not on making moves in some metagame). But does it follow from this that what I experience as the bishop checking the king must also really be – 'really' because in some sense independently of anyone's attitudes – a bishop checking the king (that it must really count as a move in the metagame)? To see that it does not, consider the following case: I see a chess arrangement of what I take to be a bishop checking the king, but I miss that there is a pawn in front of the king – and let us imagine I miss it because the 'pawn' is in fact a piece of paper that the two players playing the game are using as a pawn. If one does not like the fact that the example involves an observer observing a game, then consider the same case but with me playing the game, taking this to be a bishop checking the king, but neglecting to remember we (the players) had agreed to use the piece of paper as a pawn. The point of the example is that we cannot make the case completely independent of attitudes: we cannot say (for we would be taking some impossible position above the players and the attitudes they are relying on) that whatever it is that the players' attitudes are, there will also always be (on top of that, in a one-to-one relation) a rule according to which the players' attitudes are necessarily in accord with (i.e. a rule, at some higher level, according to which what the players do must be a move in that game at the higher level). There is no necessity about it: it depends what we, as the players, have taken to be relevant and important to playing the game. This does not mean that the players' attitudes will always trump what we, as an observer or describer, might take to be relevant in describing the players playing: all it means is that we cannot turn contingently exercised attitudes as to what is relevant into necessary ones (whether these be the players' attitudes or the observer's attitudes), and this is because the rules that may give content to an attitude are not applicable themselves (not applicable on top of, and independent of, being treated as applicable).

This first point appears to be the most important way in which Sellars tries to answer the question as what rule obeying behaviour consists in. Nevertheless, for the sake of completeness, let us note what the second and third points are.

The second point is sufficiently short to deserve full citation:

Let us note that it must not be supposed that in order to play a game at the level of rule obeying behaviour, one must first learn to play it at the level of mere pattern governed behaviour. As we have pointed out before, not all learning to play games can be learning to obey rules, but given that one has learned a language adequate to the purpose, one can learn to play (e.g. chess or poker directly as a mode of rule obeying behaviour). By 'a language adequate to the purpose' I mean, for example that one must be able to respond to certain pieces of cardboard as having 10 diamond-shaped spots printed on it, before one can learn to apply the rule language of poker. Learning to play a game at the rule obeying level does presuppose that the patterns and activities involved belong to the organism's repertoire of available discriminations and manipulations. (Sellars 1956, 221)

This passage does not help us to understand rule obeying behaviour. What it says is that someone can learn to play chess without having learnt the rules, adding that if they are to do that, they must have at least some relevant range of basic abilities, such as the ability to respond differentially to pieces of wood (for playing chess). In other words, it still assumes that what it means for me to play chess (or speak a language) just is to obey *the* rules by which the game (or the language) is allegedly exhaustively defined. But, of course, this presupposes what it was meant to defend, namely that language is a system of expressions the use of which is governed by rules. The point is that the relevant range of basic abilities is not solely identifiable by rules – one only assumes it is when one presupposes that the rules exhaustively define what it is to play chess or speak a language (and this is precisely what is in question).

The third and final point also appears in one paragraph:

In the third place, it should be emphasised that the phrase 'rule obeying behaviour' is not restricted in its application to behaviour in which one makes moves in a game via making moves in its rule metagame. There is a sense in which it is quite legitimate to say that Jones is obeying the rules of chess, even though he is not actually making moves in the rule language, and yet deny that Smith, who has learned to play merely at the level of pattern governed behaviour and hence is also not making moves in the metagame, is obeying rules. For there are many true subjunctive statements we could make about Jones and the rule language which we could not make about Smith. (Sellars 1956, 221)

Sellars wants us to imagine two persons, Jones and Smith, both of whom are not making moves in the rule metagame. Nevertheless, both of them not making rules in the metagame does not mean that they are equivalent: Jones is obeying rules while Smith is not, and that is because Jones has learnt something that Smith has not. But

what has Jones learnt, and how has he learnt it? Has he learnt it independently of pattern governed behaviour? Or is there something, in addition to the pattern governed behaviour, that he has learnt? But if so, what is this 'something'? Again, try as we might, we cannot find an answer given by Sellars.

Sellars's paper is a difficult one, and it is one of its merits that it sees the problems that face the account and genuinely attempts to answer them. But, as we have seen – and we have tried to proceed here as slowly and patiently as we could – it is not at all clear that Sellars provides an answer.

Perhaps the underlying problem is two-fold: first that Sellars begins with the assumption that language is a system of expressions the use of which is governed by rules; and second, that he attempts to find a perfect fit between the regularities of language use and language being a system governed by rules. In proceeding in this way, however, he notices that there are problems, for he does not want to 'reduce' following the rules to the order of patterns (the order of regularly performed behaviour). He thus tries to search for something to be added – some layer on top of the regularity. But this, try as he might, he cannot do: no matter how hard he tries to articulate that added layer, he runs into problems that pose again the same question. He asks us, in the end, to make a leap of faith with him: to say that when someone is playing a game and treating a bishop to be checking the king, that he really is making a move in the metagame (according to which it cannot be doubted that that really counts as a bishop checking the king).

Sellars is right to sense that the normative cannot be regularised, but his mistake is to try to make the presupposed rule-structure work by asserting that someone's behaviour is at once pattern governed and normative. The point is this: the normative is not merely regular (on this view of what the merely regular is), but for it not to be, it cannot be just something on top of the regular – it must be something inherently not ordered and patterned in the way the regular is (on this view, imagined to be). It is not an accident that Sellars cannot provide an answer as to what rule obeying behaviour is, and this because: 1) the normative cannot be explained solely on the basis of how we follow rules; and 2) the normative is not so ordered so as to be both rule following and pattern governed.

We will be returning to Sellars-inspired accounts when we look at issues raised by rule following in Part III (Chapter 10). Let us, now continue in our discussion of language as a game, but by reference to the work of Chomsky.

B. Chomsky's Rules and Regularities

To help us introduce Chomsky's problematic, let us pause for a moment and consider why and how language provides such ample opportunity for thinking that it must be (to use Sellars's phrase) a system of expressions the use of which is governed by rules (i.e. a kind of game, on one understanding of games). One answer would be that at first blush, language seems to offer an endless reservoir of examples where what is appropriate or not (what is a correct and an incorrect use) is determined – and, furthermore, where there is a domain of phenomena – grammatical rules – that seem to do the job of determining what is appropriate and inappropriate. Put differently, language seems to a domain where nothing could be more obvious than that there are straightforward cases of saying things correctly, and straightforward cases of saying things incorrectly, and that what explains these things being correct and incorrect are the rules of grammar. This may certainly be one of the reasons why language (including specific examples of allegedly clear-cut cases of correct and incorrect usage) is often drawn on to motivate the idea that there are some domains where rules (or other determinants) are clearly necessary and sufficient for making sense of appropriateness and inappropriateness. Putting things this way, however, raises an interesting question: why assume that what matters when we are trying to understand language is the correctness or incorrectness of the use made of expressions? Is it the case, say, that when we are communicating with each other, we are always paying attention (and keeping tabs) as to whether or not the other speaks correctly (in the sense of complying with a certain set of rules or principles of grammar)? Further, is it the case that even when we are evaluating someone's facility with language that what is most important is whether they are acting in accordance with what we might usefully take, for various purposes, to be the rules and principles of grammar? Or could it be more the case that we use language for a whole variety of purposes in interacting with others, and for many of those of purposes it does not matter whether the person speaks in accordance with the rules and principles of grammar? Could not treating language as if it were governed by rules and principles of grammar have its own uses, and yet not be generalisable as a way of understanding how we learn and speak a language? Why would we want to so generalise the importance of grammar?

The reason why it might be worthwhile keeping these questions in the back of one's mind is that the entire way in which Chomsky's theory of language is set up is

as a dilemma about how we come to acquire (or better how we come to have) knowledge of the rules and principles of grammar. This is a dilemma for Chomsky because if we need knowledge of rules and principles of grammar to speak a language it is difficult to see how children can have the ability to speak at such a young age – do they not need more time to learn all those rules and principles? Chomsky's famous answer appeals to universal grammar (where our knowledge of those rules and principles is said to be innate), and we shall have a look at this in a moment. But notice that such a view only makes sense if the target of explanation is how to speak correctly / incorrectly – where this is understood to be a matter of acting in accordance with certain rules and principles that are thought to determine what it is to be correct / incorrect – for it is only then that it becomes important to investigate how persons come to acquire / have knowledge of those rules and principles. If, instead, one's focus was on how persons come to be able to do things with language, including doing things with others, then it is not so obvious that the target ought to be explaining how persons acquire / have knowledge of rules and principles.

Let us animate this a bit with an example. When one isolates an expression – for instance, by having someone say to us 'the child seems sleeping' (an example used by Chomsky 1996 [1988], 560) – one immediately tries to make most salient the issue of whether the statement is correct or incorrect (or, in Chomsky's case, whether, because it is a violation of a grammatical rule,[65] it deserves, logically, to belong to the language at all).[66] In isolating this single expression, one describes the case that is of interest as one completely bereft of the context in which the expression is uttered, and thus bereft of a description that might assist us to see that for (say) the two persons interacting (e.g. a husband coming home from work, and being told the above by his wife), it matters not an iota whether this sentence is correct (in the sense of in compliance with some rule), but whether it succeeds in getting the job done (whatever that job may be – e.g. warning the husband not to be too loud).[67] The way Chomsky

---

[65] Though not said to be a heinous one – Chomsky refers to it as a 'semi-grammatical' use: Chomsky 1996, 560.

[66] Recall that on the constitutive argument one can only speak the language if one obeys the rules that allegedly govern the language.

[67] One thing that has to be avoided in such cases – though we shall not dwell on this here, for it would take us into details of the pragmatics of communication – is using the notion of a context to over-specify what would count as success. In other words, in speaking of pragmatic contexts one should not think that this is a matter of the context determining the meaning of the sentence, but something looser, e.g. seeking to influence the other's behaviour, where there could be various ways in which the

sets things up, as is visible by what he takes to be the biggest dilemma for his theory (the question of how we have / acquire knowledge of grammatical rules and principles), it is as if all that ever mattered to us when we used language was whether we did so correctly / incorrectly (i.e. again, on this view, in accordance or in violation of the rules and principles).

Having said this, it would be too quick and too harsh on Chomsky to say that he ignores the variety of abilities and skills we use when we employ language. For instance, Chomsky does sometimes speak of 'pragmatic competence' (distinguishing it from 'grammatical competence'), but it is telling that when he does so, he characterises it by reference to its role in relation to the knowledge of rules and principles. Thus, Chomsky says, 'pragmatic competence underlies the ability to use such knowledge along with the conceptual system to achieve certain ends or purposes', and then he also adds: 'It might be that pragmatic competence is characterised by a certain system of constitutive rules represented in the mind, as has been suggested in a number of studies' (Chomsky 1980, 7). Chomsky does not enumerate here the examples of 'pragmatic competence'; instead, he suggests it is definable in relation to 'grammatical competence', which he says consists in 'the cognitive state that encompasses all those aspects of form and meaning and their relation, including underlying structures that enter into that relation, which are properly assigned to the specific subsystem of the human mind that relates representation of form and meaning' (Chomsky 1980, 7). Even, then, where Chomsky does speak of other abilities and skills being important, they are important always because they either support or are involved in the acquisition or manipulation of the knowledge of rules and principles.

For Chomsky, then, language understood as governed by those rules and principles (and thus by determinants of appropriateness and inappropriateness) always and necessarily serves as criteria for what matters about understanding language and indeed whether any particular account of language is a good one. In dealing with an objection to his account – one to which we will return but one which, in brief, is based on the idea that perhaps some of our abilities and skills to speak a language may be more like 'knowing how to ride a bike' than knowing that something is appropriate or not (with respect to some rule) – Chomsky says: 'I know of no other account' (than

---

response by the other could be appreciated as appropriate (e.g. the husband could leave the house and go outside for a drink, or he could stay but whisper).

his own, or versions of his own) 'that even attempts to deal with the fact that our judgements and behaviour accord with and are in part explained by certain rule systems (or to be more accurate, are explained in part by theories that attribute mental representations of rule systems)' (Chomsky 1980, 12). For Chomsky, the task of the theorist of language (including, or perhaps especially the critic of his approach) is to provide an account 'of how it is that what speakers do is in accordance with certain rules, or is described by these rules… If', he continues, 'someone can offer such an account of how it is that we know what we do know, say, about reciprocals, or judge as we judge, and so on, there will be something to discuss' (Chomsky 1980, 13). Here, Chomsky is effectively saying that one cannot criticise his system without explaining what his system seeks to explain – but what he fails to see is that where his system goes wrong is in what it seeks to explain. Chomsky's work is contestable not at the level of how to operationalise some system of rules, but rather in thinking that that is what needs to be operationalised.

Notice that none of the above is dismissive of the importance of the rules and principles of grammar. It is certainly true that rules and principles of grammar are important, and it is vital that we do not ignore the various roles they may play in both understanding what it takes to speak and also speaking language. For instance, anyone who sought to explain our use of language but failed to consider how we do sometimes (especially in pedagogical contexts) criticise one another for saying something and do so on the basis that it is a violation of some grammatical rule or principle would be missing an important dimension of language use, including how one may come to speak language successfully in a community. Of course, someone would also be missing a great deal about how language may be learnt if they neglected those many cases (including in pedagogical contexts) where a teacher does not and perhaps could not cite (and perhaps could not because there is no) rule or principle according to which some usage is correct or incorrect.[68] The point here is that Chomsky's object is too narrow: it is not rules and principles themselves that ought to somehow delimit what we ought to be explaining. At the same time, however, in broadening out the object of explanation (when making sense of

---

[68] For example, consider the usage of 'that' and 'which' in English – here, we are much more likely to feel our way about often to such an extent that where some attempt is made to introduce a rule (as in Microsoft word, which operates on the rule that all 'which's' must come after commas, and all 'that's' cannot be preceded by commas) we find such a rule much too rigid and, indeed, better not followed.

language) we ought not exclude rules and principles, but instead look carefully at how and when we make use of them.[69]

It was mentioned above that the object of Chomsky's theory is to explain how the rules and principles of grammar are operationalised, i.e. how persons have them and come to act in accordance with them. Notice, in this, the one-on-one mapping of linguistic correctness (determined by the rules and principles of grammar) and the regular performance of that which is required by the rules and principles. Chomsky argues that we do use language (on the whole) in a certain way, and that the way we use it is (again, on the whole correctly), i.e. in accordance with the rules and principles that determine what is correct, and that the task of the theorist is to explain how we have / acquire knowledge of such rules and principles, it being assumed that it is this knowledge that we draw on when we speak correctly on any one occasion (e.g. we apply the rules to new expressions). There is an obvious question to ask at this point: how has Chomsky identified the regularities of language use? Has he not done so on the basis of the rules and principles – indeed, are not the rules and principles generalisations from language use? So in what sense can the object of explanation be how the rules and principles of language are operationalised, e.g. how they are learnt and acted in accordance with? Is it not that the rules and principles are already presumed to be regularities? This is a question that will accompany us throughout our discussion below.

We can put this question another way: how can the theorist be sure that the rules she has identified – that is, the rules and principles of grammar – are also the rules that the persons who are described as acting in accordance with have learnt and are following? Without being able to show that the rules and principles of grammar identified by the theorist are also the ones that are being learnt and followed, it seems difficult – if not impossible – to distinguish the theorist's list of rules and principles from other such lists. Chomsky defends himself here in two principal ways: first, by defending the reality of rule following; and second, by challenging the very notion of learning (saying, as we shall see in a moment, that we do not learn the rules, but that the rules 'grow in us').

---

[69] Again, the point is not to regularise this 'use', i.e. not to make it necessarily the case that the way rules are used constitutes a pattern or order, defeasibly applied even when we do not realise it. The relevance of the rules must remain contingent.

Note that the defence of the reality of rule following cannot merely be that a person can, on some occasions, represent some rule to themselves and, as a result, act in accordance with it. It has to be that persons on the whole follow that particular rule. 'Follow' has to involve some kind of mental process, and thus the difficulty that Chomsky has is to show that every time we act in accordance with the rule, we also follow it (recall, again, that if he does not show that, he will not be able to claim any special status for those rules – they will just be one out of many generalisations). Given that whenever we speak, on Chomsky's account, we (on the whole – i.e. unless we make a mistake) act in accordance with the rules, this means 1) that we do a lot of rule following; and 2) that there are a lot of rules. The challenge, then, is to answer queries about how we could be engaged in so many mental processes every time we speak language (and especially so when some of the persons in question may be the children and some of the rules could be quite complex), and answer queries about how we can be acting in accord and following rules when there are disagreements over what those rules are, when new grammatical rules are being proposed often, when there seems to be quite a lot of diversity of rules in different communities (even though those communities can otherwise communicate successfully to each other – one thinks of dialects / slangs etc).

Chomsky certainly gives us reason to think that sometimes persons can represent rules to themselves such that we could say that it is that representation that is the cause of the action – but it is difficult to see where he gives any argument that such representation, by those very rules, is necessary; it seems that it is only necessary if one's purpose is to try to show that the regularities of language use one has picked out, and identified by reference to such-and-such rules and principles of grammar, are the very the same ones that persons represent to themselves. But this is precisely at issue: the way the theorist describes the regularities of language use (calling them rules and principles of grammar) simply need not be the way persons learn to, and then do, speak the language.

Chomsky seems, at times, to acknowledge that he has no argument for the necessity of representation by the very same rules as the rules and principles of grammar (in effect, regularities of use) picked out by the theorist. Instead, he appeals to this being possible. For instance, in his reply to the 'know-how-bicycle-critic', Chomsky argues that although it is possible that the bicycle rider does not represent rules to himself, such that this process is what explains him being able to ride a bike,

it is also equally possible that he is: 'suppose…in fact the bicycle rider does have a representation of certain physical principles in his mind and uses them to plan or compute his next act. In this case we should attribute to him a cognitive structure, and in fact, it would be quite appropriate to say that he cognises these principles as he does the rules of his language. The question I take it, is basically one of fact' (Chomsky 1980, 11). But Chomsky cannot simply establish that it is possible – he must show that it is these, and only these, rules (i.e. precisely the one he has identified as the rules and principles of grammar) that the subject represents to herself. Perhaps it is conceivable to imagine instances where some bicycle rider (perhaps a physicist) will represent certain physical principles to themselves when riding a bike – but it is hardly conceivable (and thus hardly a matter of fact) that a description of bike riding in terms of physical principles must coincide with an explanation of how persons, every time they ride bikes, produce the action of bike riding by representing those principles to themselves. Even the physicist bike-rider, who represents some principle to himself as he rides, will presumably not be producing all of his actions on the basis of that principle – does Chomsky mean to suggest that the rider is somehow processing all of the possibly relevant physical principles at the same time? And what about all the physical principles we have not formulated (that we do not think, as yet, are involved in knowing how to ride a bike) – is the bike rider 'representing' these to himself as well? Chomsky chooses a curious example to 'help clarify the issue' (Chomsky 1980, 11). It is worth setting this example out in full:

> Consider two missile systems, each of which is designed to send a rocket to the moon. One of them operates along lines once proposed by B.F. Skinner; it has several pigeons looking at a screen that depicts what lies directly ahead, trained to peck when the rocket veers off course, their pecking restoring the image of the moon to a focused position on the screen. Consider, in contrast, a system that incorporates an explicit theory of the motions of the heavenly bodies and information about its initial position and velocity and carries out measurements and computations using its internalised theory to adjust its course as it proceeds. This rocket might hit the very same spot as the servomechanism with the pigeons, but it would do so in a very different way. Mere investigation of behaviour might tell us little, perhaps nothing. A deeper look might be required to distinguish the two systems. In the second case, but not the first, inquiry might lead us to attribute to the missile something like a 'mental state'. That is, it might lead us to formulate an abstract characterisation of perhaps unknown mechanisms, postulating a system that involves the cognising of certain principles and representations. In the first case, such an account would be factually wrong. I think that the two cases fall

on opposite sides of an important ivied, and that the second – the cognising missile – shares properties with human knowledge. It also lacks crucial properties; for example, it is a task-oriented device, whereas knowledge of language, for example, is not. (Chomsky 1980, 11)

This is an instructive image, for it shows how fixed Chomsky's picture of language is. What we ought to do – how we ought to speak – and also the target of explanation (as is the target of the missile), is given and identified in advance. In the case of language, this target is itself described in terms of rules and principles, but only in the sense of persons acting in accordance with those rules and principles (on the whole). What Chomsky is trying to set up here, by way of analogy, is a way in which we can see how persons could be like the second missiles: getting to the target by some kind of cognising process involving the representation of those rules and principles. There are two ways to read this: one is to say that Chomsky has shown that it is possible to conceive of the production of a certain kind of action (already identified as the action to be performed) in terms of a certain means, but that he has not shown that those means must always be the means the theorist has identified (again, recall that he needs to show this – otherwise we can offer many descriptions of the means, none necessarily of greater stature than the other). The other is to say that Chomsky has shown that although the theorist may not have shown that the means the subject uses to get to the target is the means that the theorist has identified, the theorist is within the ballpark and it is just a matter of getting the description of the means more and more accurate until one will achieve a perfect fit (it being the point that a perfect fit is conceivable). Chomsky would presumably prefer to say the second thing – but is this answer really open to him when, on his account, the target has already been identified by the theorist and described in terms of conformity with the rules and principles? In other words, the target in the case of language seems to quite different from a missile starting out from one place, and hitting another place, where in the case of the missile this starting position and the target are not themselves described on the basis of the same rules and principles that it is suggested are also the means for getting there.

Return for a moment to the last sentence in the quote above: that knowledge of language is independent from being task-oriented. At various points in his work, Chomsky gives examples of persons who have knowledge of language, but cannot exercise it, i.e. put differently, his account of why some persons do not have the ability to speak language is not that they do not have the knowledge (for that

knowledge is, on Chomsky's account, in large part innate and consists in knowledge of the rules and principles of universal grammar), but that they are suffering from some obstacle in exercising it. The image of knowledge here is that of an isolated brain cognising an abstract system of rules and principles – perhaps quite like a computer, or a missile. As Chomsky puts it himself, 'To know a language…is to be in a certain mental state, which persists as a relatively steady component… To be in such a mental state is to have a certain mental structure consisting of a system of rules and principles that generate and relate mental representations of various types… If…to know a language is to be in a certain mental state comprised of a structure of rules and principles…then in theory one could know a language without having the capacity to use it' (Chomsky 1980, 5). I might lack this capacity because of difficulties with memory or attention (Chomsky 1980, 6).

This returns us to the distinction above between 'pragmatic competence' and 'grammatical competence', such that we can make the same point by reference to those terms, namely 'that it is possible in principle for a person to have full grammatical competence and no pragmatic competence, hence no ability to use a language appropriately, though its syntax and semantics are intact' (Chomsky 1980, 7). Chomsky offers an analogy (drawing on some work by Ave Kasher) with 'the case of a policeman who knows the syntax of traffic signals (red and green lights and their sequence, etc) and their semantics (red means stop, etc) but lacks the knowledge of how to use them to direct traffic' (Chomsky 1980, 7). Again, note the image here: knowledge – of the kind we ought to be interested in explaining – is knowledge of rules and principles. Ideally, this is knowledge that will be utilised, but it also may not; knowledge of this kind can stand alone.

The following pertinent question arises: how conceivable is it that one can pry apart knowledge and capacity in the way Chomsky does? Notice that one could say that persons differ in capacity to speak the language (which of course Chomsky recognises), but explain this not on the basis that those persons differ in their ability to make use of their knowledge, but that their knowledge of language and their capacity to use it are one and the same thing: they develop together. Chomsky cannot say this because he is committed to the view that much of our knowledge of language is innate (and comes in the form of the rules and principles of universal grammar). But how does this gel with the analogy of the policeman: is the policeman somehow born with rules and principles for directing traffic? Clearly, not, so we must imagine that the

policeman has been taught these rules and principles, but still does not know how to use them. First of all, this assumes that all that is involved in directing traffic is knowing how to use the rules and principles for directing traffic – and this not only stretches credulity (for there are, surely, infinite many cases not conceived of by the rules and principles that the policeman will confront; if there were not, could we not just program a robot to direct traffic?), but it also leaves without explanation the development of the ability to use rules and principles. Second of all, it requires Chomsky to posit that persons learn – in some classroom – all the required rules and principles of the language, for otherwise the analogy will break down (we have already said that it is inconceivable that the policeman is born with knowledge of the rules and principles for directing traffic). But, of course, given his commitment to universal grammar, and his image of knowledge, Chomsky would not want to say that we learn rules and principles that way.

We shall see in a moment how this view of knowledge affects Chomsky's account of learning. But first, recall that the reality of rule following not only had to deal with the difficulties of requiring persons, every time they speak, to represent to themselves rules and principles, but also with difficulties concerning how many rules and principles there must be, how there can be so many different ones, how we could create them (including create them for specific purposes, such as writing poetry), as well how we could have disagreements about what they are. It is not easy to pin Chomsky down on this issue. At times, Chomsky suggests that some knowledge of rules and principles that children have they acquire through inductive generalisation, but then he also stresses that 'much of our knowledge reflects our modes of cognition and is therefore not limited to inductive generalisation of experience, let alone any training that we may have received' (Chomsky 1980, 4), and he offers the following image in support of that last point: 'And just as the visual system of a cat, though modified by experience, will never be that of a bee or a frog, so the human language faculty will develop only one of the human languages, a narrowly constrained set' (Chomsky 1980, 4).

Again, it is here assumed that there is such a thing as 'one of the human languages', and that this thing is a system of rules and principles, and it is to one of these moulds that we must fit any person's capacity to speak language. There is something arresting about the analogy of speaking language with cats' vision, but is it one that needs to depend on their being 'systems', either of language or of vision?

179

One certainly would want to say that cats have something (some biological make-up) that allows them to develop abilities in certain ways (that are not identical to our abilities to see, or to those of frogs or bees), and thus one would not want to say that those abilities that cats develop to see come from nowhere – but must one, in order to acknowledge this, go to the other extreme and say that whatever is developed by cats, bees, frogs or us, in case of our abilities to speak language, can only be explained on the basis that it somehow 'reflects', or is nothing more than an instantiation of, what we had at birth (i.e. some kind of knowledge of rules and principles)?

The problem here with Chomsky's view is that anything can be said to 'reflect' some allegedly more basic rule and principle. As soon as the theorist discovers some new way of articulating what counts as a certain patterned use in language, she can also argue that this reflects a certain more basic rule and principle that she is now making explicit. Further, and connected to this, what is missed (and presumably this is part of Pinker's disagreement with Chomsky) is how the allegedly more basic rules and principles – or at least that which is carried over from generation to generation, and comes innate – is subject to change by what is experienced, what is developed, by the next generation. Not everything, surely, is always and necessarily merely a 'reflection' or instantiation of what was, for 'what was' – including what is passed on to the next generation – changes, as the species evolves. In answering any such worry, Chomsky would hasten to add that he speaks of the 'interplay of genetically determined principles and a course of experience' (Chomsky 1980, 13), but it is difficult to mesh this general statement with the way that Chomsky discusses knowledge, especially the way in which (as we have seen above) he constructs the target to be explained and argues that all theories of language must meet it – the point is that the very construction of that target is one that does not leave room for a more dynamic account of the development of the ability to speak (which must, surely, include a concept of knowledge that is integrated with the ability to use it, including as part of various, indeed innumerable, kinds of tasks).[70]

To complete this discussion of Chomsky, let us briefly tackle the question of what learning becomes on this view of the knowledge of language. What is instructive, and perhaps unsurprising given some of the comments made above, is that

---

[70] Again, the challenge here is to avoid describing a catalogue of pre-given tasks, for if one did do that one would risk producing another version of Chomsky's view, but at the pragmatic level, i.e. one would see us as pre-programmed to perform certain tasks, where any 'new' task could be described as but an instantiation of a more general (more basic) task.

Chomsky questions the viability of the concept of learning. He says, for instance, 'I would like to suggest that in certain fundamental respects we do not really learn language; rather, grammar grows in the mind' (Chomsky 1980, 13). Of course, part of the controversy here will be what is meant by 'learning'; for Chomsky, this includes learning processes such as 'association, induction, conditioning, hypothesis-formation and confirmation, abstraction and generalisation, and so on' (Chomsky 1980, 13). Insofar as learning is said to consist of such processes, he says, 'it may well be that language is not learned' (Chomsky 1980, 13). In order to make this more plausible, Chomsky provides us with an image of 'the heart, or the visual system, or other organs of the body'; when we speak of these developing 'to their mature form, we speak of growth rather than of learning' (Chomsky 1980, 13). As we saw above, Chomsky is here assuming that there is a 'mature form' to be reached – that, as he puts it, there is a 'final structure' to be 'attained' (Chomsky 1980, 13). Once again, and even in the context of learning, the image is one of a system of rules and principles to which all must conform, if one is to count as speaking a language – but with the twist that somehow this structure is already inside of one, such that achieving a developed form means, perhaps, not suffering from any obstacles that allow one to show that one does attain (or better, realise or exemplify) this structure.

Chomsky softens this somewhat when he says that he does not mean to 'demean the content of what is learned', but then he adds, somewhat remarkably, that 'what is significant for human life is not necessarily significant for the person inquiring into human nature' (Chomsky 1980, 14), which suggests that science not only is, but ought to be, so autonomous as to neglect (and treat as 'merely' significant for human life) all that which is not already explicable by the rules and principles it identifies as governing the usage of language. All that falls outside of the realm of those rules and principles (including being characterisable as 'reflecting' allegedly more basic, more universal, rules and principles) is said to be nothing more than an 'idiosyncrasy' (Chomsky 1980, 14) – though, Chomsky adds, 'even here it is highly likely that powerful intrinsic constraints guide the course of development' (Chomsky 1980, 14).

We have not here inspected many features of Chomsky's view, and even with what we have looked at, there is scope for a much more fine-grained analysis. But it is hoped that sufficient indication has been given of the view – sufficient at least to see how it implausibly prioritises the idea of some domain (in this case language) being

necessarily and sufficiently governed by some rules (or other determinants, such as principles that allow the instantiation of more specific rules, of what is appropriate and inappropriate). None of the criticism above ought to make us think that this means we need to abandon the importance of rules and principles, including the role of grammar in our learning and exercising the capacity to speak language, as well as our participation in communities of speakers where we might have apportion praise or blame at least on the ostensible basis of conformity with or violation of some rule or principle of grammar. But we need to loosen the hold of the idea that language is a system of expressions, where what is a correct and incorrect usage it is determined by some rules and principles, such that all we do when we speak language, and all we need to do in order to speak it, is acquire and exercise knowledge of those rules and principles. To do this, we not only need to question the notions that appropriateness / inappropriateness require being determined and that there are things that do that job, and not only that language is an activity the nature of which (as with games, on some views) is or can be determined by certain determinants, but also certain other notions, including the idea that what matters most when we speak language is whether we speak it correctly, or that what matters most is whether we have a certain knowledge of the language (rather than, say, an ability to communicate successfully, or to perform all kinds of task where speaking a language plays a role).

Most importantly, perhaps, we need to resist the idea of a target or object of explanation that is already conceptualised as a matter of acting in accordance with certain rules or principles, and thus where everything (at least everything 'significant') is understood to be nothing but a 'reflection' of those rules and principles. To create a target or object of explanation that makes acting in accordance with rules and principles is akin to imposing one fixed order onto another: the order of regularity (repeating the same patterns) and the order of the normative (imagined as a system of rules and principles). This mapping of one onto the other is also the source of Chomsky's troubles concerning his picture of the mind learning and working (i.e. he has to depict it, implausibly, as already in possession of the rules and principles and as ceaselessly representing them whenever speaking). The alternative is to see that the rules and principles of grammar are contingently useful resources (especially when, as in a context of justification, there may be reason to claim grammatical correctness), and then to focus on our pragmatic abilities to use language to do an indeterminate range of things. In essence, we over-generalise the relevance and

importance of rules and principles when we create a target or object of explanation in which what we are said to act in accordance with is also what we follow. When, instead, we picture rules and principles as contingently useful resources, we then allow for guidance by rules and principles, but we recognise that there is danger in understanding the normative to be reducible to a re-enacted order, and danger in thereby characterising ourselves as so rigid and passive. As we shall see later, retaining the sense in which we are active and creative vis-à-vis that which can be experienced as appropriate and inappropriate, and thus rescuing our ability to be hesitant and tentative, is of vital importance to interaction understood from the second-person perspective.

C. Against Language as a Rule-Governed Game

It is time now to look briefly at some of the literature that has resisted this image of language as a system of expressions the use of which is governed by rules (and thus an image that would have us think that language can be understood as a game, on a certain understanding of what a game is). Having done so, we will finally be in a position to summarise our discussion of the constitutive strategy.

There is a curious thing one notices in some of the literature that resists the analogy between chess and language: it saves language at the cost of sacrificing chess. Put less colourfully, it argues that language is not like chess because, unlike chess, its nature is not determined by rules. A perfect example of this is Douglas McGee's paper, 'Fun, Games and Natural Language' (1964). The 'rules of chess', says McGee, 'are perfectly explicit, finished and closed – have been in general since about the sixteenth century, and are in particular for any given match' (McGee 1964, 337). The 'rules of chess can be broken, but never bent' (McGee 1964, 338). Chess 'is a closed universe, as a whole it has no environment' (McGee 1964, 338). 'All you ever need to know' to play chess 'is the rules: in explaining, understanding, or playing a game, one can as well consider that the whole of it was created, *ex nihilo*, a few minutes before one started to play' (McGee 1964, 338). 'Knowing the uses of chess-men is much like knowing how to operate with a calculus' (McGee 1964, 341). 'By reference to the rules of chess one can without obscurity or argument refuse some moves as "illegal", "impermissible", or "not correct"' (McGee 1964, 341). Chess is 'neat and clean…free of factual and of metaphysical taint…I can offer nothing so sterile and cutting' (McGee 1964, 341). The rules of chess are 'fixed, detached,

deontological' (McGee 1964, 343). These are fighting words, indeed! It is difficult to find a more ruthlessly deterministic view of chess.

The view is all the more strange in that McGee has (as we shall see in a moment) a very different understanding of language, which can still be said to some extent to be governed by rules, but rules that work differently (so it is not his conception of rules in general that is at issue). Further, he understands appropriateness and inappropriateness to work differently in language – so, again, we are not dealing here with assumptions made about what is necessary for appropriateness and inappropriateness. Finally, at the beginning of the paper, McGee points out that it is possible to speak of playing without playing a game – which suggests he might have a more flexible image of games. He says, in this respect, referring to Wittgenstein's use of 'Spiele', that in German the term 'Spiele' can refer to 'some pretty unstructured activities… It is possible to play without playing a "game" – think of playing cars, or Cops and Robbers, or just playing – but since in English we cannot say "speilen ein Spiel", but must say "playing a game", the notion of games and the suggestion of rules have come into English-thinking philosophy through a crack in the English language' (McGee 1964, 336, fn. 1). Thus, McGee is aware that it may only be a peculiar feature of (English) language use that connects games in some exhaustive way to rules, but he nevertheless insists at length that our understanding of chess ought to be immune to any such worry about its relation to rules.

We shall not repeat here the arguments made in the preceding Chapter about an understanding of chess, and other games, based on the alleged exhaustiveness of rules – just briefly, let us recall that we gave plenty of examples about where we might say someone was playing chess but not following the rules, and where we might say that someone was not playing chess despite following the rules. It is only by neglecting such counter-examples that we can so over-generalise the contingent relevance of rules so as to construct an image of chess as a game exhaustively determined by a certain defined set of rules. But what is of interest here is not so much the image of games that McGee uses, but why he does not think it applies to language.

McGee argues that 'the rules of a natural language are at once less distinctively fixed and more clearly changeable, and less explicit and more flexible than are the rules of chess', e.g. 'the uses we make of language – coping in one way and another with the physical and social work – in their success and failure react

upon, condition and modify rules of use' – indeed, McGee adds, 'to say this is really to say little more than that language is learned and has its uses in an extra-linguistic environment' (McGee 1964, 338). Further, linguistic rules do not have the kind of 'independence or priority' that chess rules do, i.e. in chess, 'the uses permitted to chess-men depend on autonomous and antecedent rules…[which] leads us to think of rules of use as prior to all uses, and independent of usage', this being untenable with respect to language, for 'we shall have to say' (insofar as we talk about linguistic rules at all) 'that they are determined by usage' (McGee 1964, 340). 'Successful use', says McGee, in the context of language, 'depends on time and knowledge of circumstance, on estimates of appropriateness and possible success, on the likelihood of one permitted response and the incongruity of others… knowledge of facts, of use, usage and rules is in principle interdependent, and there is no fixed or clear order of dependence or priority among them' (McGee 1964, 341). Let us pause here for a moment, and consider these ideas.

Imagine we (i.e. students and a teacher) are in a classroom and following a textbook with rules of grammar in them. The teacher asks students to do a certain exercise: the students use the rules to guide them, and the teacher uses the rules to evaluate them. Here, we might want to say that the rules are somewhat independent of usage: after all, it is not whatever the students say that the teacher will praise; rather, the students' usage will be subject to the rules in the textbook.[71] The point here is to show that rules of grammar can be used in a very similar way to the rules of chess: as somewhat independent of usage, and as prior to or antecedent to usage. It is not that it is somehow not part of the life of language for rules to be used this way: all it means is that rules can be used this way. Similarly with the rules of chess: rules can be used to guide persons (especially when they are first learning – though it is possible to learn chess without learning the rules) and they can be used to evaluate behaviour of playing chess (though they also may not, as when we play in a pub and it need not matter whether everyone knows the *en passant* rule, or everyone thinks that when one promotes a pawn one must take a Queen). Of course, there are some differences, e.g. it is true that we tend to agree more readily on what the rules of chess are and the

---

[71] Nor do these rules have to reflect the teacher's linguistic abilities (they may simply be effective pedagogical instruments).

circumstances in which they tend to be relevant are fairly well established.[72] Further, disagreements about what the rules are, or what their scope is, is relatively rare (but certainly not non-existent, and thus certainly not impossible – in this respect, the very fact that the rules of chess changed in the sixteenth century is an example).

Thus, whether we speak of chess or language, we can use rules in all kinds of ways, but when McGee refers to 'usage' he seems to have something else in mind, i.e. our acting in accordance with rules that describe patterns of use. This is where we get into difficulties, for we attempt a perfect fit between rules-as-regularities and rules-as-guides. Insofar, then, as McGee sets up the alternative image of language (alternative to it being a system governed by rules like a game) as one of a patterned order of rules-as-regularities that, in addition, persons are said to follow, then that image is one that we must resist. It is, in effect, just another version, and not an alternative, to the view of language as a game (where games are understood as domains governed by rules).

The closer one looks at McGee's paper, the more one sees how the normative and the regular overlap in his account. Thus, for instance, McGee argues that one of the distinguishing marks between chess and language is that the latter, unlike the former, is essentially a communal activity (McGee 1964, 342). But what McGee finds, as soon as he begins to speak of a community, is that he is drawn to think of it as constituted by 'behavioural regularities recognised by the group as norms' (McGee 1964, 342). McGee here banishes rules to 'explicit rules', and begins to speak not only of norms, but also of implicit customs (McGee 1964, 343). There could be no clearer statement of the regularisation of the normative: of reference not only to behavioural regularities, but behavioural regularities that are sometimes recognised as norms (or, when not recognised, then said to be awaiting recognition, as if they were implicit norms). What we want to say instead is that we should not banish the rules (conceived of as explicit): they are contingently useful normative resources. Further, when it comes to behavioural regularities, customs or norms we need to see these as ways of treating social life as patterned – and we need to see that the content of any such treating (the specification of the pattern) is an explanation that cannot be cashed out in terms of those regularities, customs or norms being somehow internalised. Such

---

[72] Though their being well established does not mean that they are relevant by default: their relevance remains contingent. The point is that the rules are never sufficient by default – it is just that one can treat them, contingently, as sufficient to describe and / or evaluate what is happening.

explanations may be better or worse, but they cannot be elevated to a status of necessary, even by defeasible default, guides relied on by persons when they act and interact.[73]

The contrast McGee draws between chess and language is much too stark. We are not talking, when speaking of chess and language, about two different things: a closed universe (of explicit rules) versus a less closed one (to recall, McGee's description of language is not less closed, for it makes it exhaustively dependent on regularities recognised or potentially recognisable by a community as norms). Instead, we are talking about attitudes we can take to activities, where such attitudes may 1) make contingent use of normative resources to treat something as required, and unchangeably so, or 2) refer to some specification of the way in which one takes the activity to be ordered or patterned (referring to customs, etc, where what makes these non-normative is precisely the absence of any mediating or critical reflective attitude). The point is that one can take either attitude to chess or language. The mistake we make is where we depict chess or language in such a way that what is appropriate to do (to play chess or speak a language) is exhausted, in advance, by an order of rules or norms (in short, by determinants of appropriateness and inappropriateness). This is a mistake because it over-generalises the contingent utility of rules or norms.

Even theorists who argue explicitly against understanding language as a game (where games are understood in a certain way) come, then, to prioritise talk of rules or norms. Ganz, for instance, whose work we encountered in Chapter 3, raises several problems for a view (which he attributes to Chomsky, amongst others) according to which we follow the rules of grammar when it is impossible (or at least implausible) to think that any one of us can know all the rules there are to know. The level at which Ganz raises difficulties is at the level of being guided or directed by a rule: he argues that it is nothing but the hope of a theorist 'that the structure revealed in language (and set down as grammar) is like the internal structure of speakers' (Ganz 1971, 110). This is an important observation that we have also made above, but when Ganz comes to offer an alternative, this is what he says: the object that needs explaining when it comes to linguistic ability is 'the ability to be grammatical' (Ganz 1971, 107). There is an important ambiguity in Ganz's elaboration of this ability. On the one hand, he says this: by this ability 'shall be meant the ability to recognise

---

[73] Another way to say this might be to say that guidance – one of the attitudes one can see as part of the normative – is not regularisable. It is, instead, at least sometimes, tentative, hesitant, and exploratory.

"new" sentences as grammatical or not and the ability to produce "new" grammatical sentences without knowing, i.e. being able to articulate, the grammatical rules for the language' (Ganz 1971, 107). What is key here is whether Ganz thinks that what is 'grammatical' is always and already presented to us – whether it is, as it were, already grammatically correct or incorrect by virtue of some rule, or whether we actively and creatively construe what we might go on to treat as grammatical. Ganz's discussion is brief so one cannot give a definitive answer, but some signs of how he would treat the topic can be found in his citing, with approval, the following statement by F.A. Hayek: 'The most striking instance of the phenomenon from which we shall start is the ability of small children to use language in accordance with the rules of grammar and idiom of which they are wholly unaware' (quoted in Ganz 1971, 117). Ganz wants to avoid the problems that are faced by those theories that try to account both for regularities of language use (in the form, say, of rules) and for those rules being processed internally by persons. But what he does is let go of the latter, while keeping the former, and this is no solution – indeed, the very problem is with the prior identification of certain patterns of language use to be thereafter, passively and uncreatively, re-enacted or recognised by persons (and thus describing linguistic ability as but pattern recognition, which is what Ganz essentially does: see Ganz 1971, 122).

What we see here is that the governance view works not only by reference to making conventions, norms and rules necessary, but by prioritising determinants of appropriateness and inappropriateness (or grammaticality / ungrammaticality) in general. The problem, then, is the presupposition of an order – ontologically secure, and the epistemological target. It is this that makes us describe persons as passive and uncreative. That makes it more difficult to see how we can be active and creative, and yet non-assertive and non-committal – but tentative and hesitant – at the same time. If we want to rescue the tentativeness and hesitancy and, at once, the activity and creativity, of persons, which is vital for our ability to participate in social life from the second-person perspective, including when playing chess and speaking a language, then we must avoid seeing these as orders (or only as orders). The point is to include, in our understanding of chess or language, room for irregularity, unpredictability, uncertainty and ambiguity – a sense, in short, in which that which may be relevant (including to an appreciation of what is appropriate or inappropriate to do) has not yet emerged but may emerge in the course of playing chess or speaking language.

We have come to the end of this second Chapter on the constitutive strategy. By no means have covered the field, but we have also seen its typical manifestations, and in two contexts: games and language. In trying to rebuff the constitutive strategy, we have not drawn on many other critiques of constitutivism (e.g. see Railton 1997, and Enoch 2006). One can argue, as these theorists do, that constitutivism does not capture all that it needs to about the experience of being able to resist the pull of rules and norms (and normative resources generally),[74] or that no amount of constitutivising can determine whether we ought to do what the constitutivist wants us to do,[75] but analysing these attempts in detail is outside the scope of our discussion. In any event, we need not rely on these forms of arguing against the constitutivist: we can, instead, as we have be trying, show that it is implausible to think that the normative resources we may take sometimes to be useful in guiding ourselves and evaluating others are also by default applicable as the way in which persons, e.g. play games or use language. The basic problem with the constitutivist strategy is that it over-generalises from what is contingently useful to some kind of necessity, even if only a defeasible one. For a much stronger claim to necessity – a metaphysical, *a priori* one – we need to turn to the next Chapter's discussion of 'super-norms'.

---

[74] That its 'must' is much too strong, and ends up not being able to account for mistakes, at least when mistakes are understood on a richer model of agency: this is the brunt of Railton's 1997 critique.

[75] The constitutivist cannot guarantee that we ought to engage in belief understood to necessarily be a matter of aiming at truth, for it may be that what I ought to do is believe and aim for something else: this appears to be the basic message of Enoch 2006.

Chapter 10. Super Norms

Before we complete this Part, we need, in this final Chapter, to consider one more version of the stance of the governance view (its first element), namely that determinants of appropriateness and inappropriateness are necessary. This final version of the first element is based on the basic idea that there is some 'inherent normativity' in some domain or activity, and this inherent normativity consists in some 'super norms', i.e. norms that are always and necessarily applicable and that are not revisable. Claims like this are sometimes made about (some or all?) the norms of logic, where these norms are said to characterise the inherent normativity of rationality. Although we cannot, of course, give justice to the debate over the nature of rationality,[76] nor indeed logic, it is important that we look briefly and critically at attempts to stake out this very special status for the norms of logic. To repeat: the point, as also with the previous Chapter on language, is not to make a contribution to our understanding of logic. Instead, it is to tackle the form of argument – the reference to so-called super norms – that appears, very often, in the context of discussions of logic.

There is one matter that we ought to place clearly on the table before we begin. This is the idea – or, more often, assumption – that if we cannot show that there are rules of logic that have this special status of super norms, then it follows that rationality is conventional, in the sense that it is determined by social norms (that are arbitrary – could have been otherwise – in a way that super norms are not). This divide is a false one: both sides of the divide are examples of the governance view, for both think that what is rational has to be determined; they just differ in the source of those determinants (*a priori* or conventional). Both views are mistaken in how they picture rationality. What we need is not a picture that assumes that what is rational can and ought to be determined (by determinants), whatever their source, but rather a picture of the exercise of reason (an appreciation of reasonableness, we might say) that recognises that part (but not an essential, and not even an exhaustive part) of that exercise may involve occasionally treating certain normative resources as applicable.

---

[76] For instance, we will be ignoring the debate surrounding Broome and Kolodny's disagreement over whether rationality itself is normative, which also means we will be ignoring some of the distinctions in this literature, such as distinctions between reasons and requirements, and between wide and narrow scope of reasons, etc (see Kolodny 2005; and Broome 2005 and 2007).

Reasonableness versus rationality, then, and not *a priori* rationality versus conventional rationality, is the relevant divide.[77]

In fact, the reference to reasonableness and rationality raises another important point that deserves mention at the outset. It has been argued by some that rationality, understood in terms of 'perfect rationality' (as an idealisation) is not the kind of understanding of rationality we should have because the rules that define perfect rationality could not possibly be the rules that guide persons in being rational.[78] This highlights the limits of accounts of perfect rationality (of course, idealisations can still be useful – but dangerous, too, if they are relied on too heavily as predictions, as the use of *homo economicus* models has often revealed), and is welcome for that reason. However, it can also run into another problem, namely it has the potential to make us place all our bets on rule guidance (just of the right, i.e. more modest, kind of rule) as the correct picture of rationality, i.e. we might come to think that all that is involved in understanding rationality is how we are guided by the right kinds of rules (these being, by necessity, not the rules that define perfect rationality). The advantage of a concept such as reasonableness is that it does not rely on a picture of rationality where we act rationality only if we are guided by the (right kind of) rule and irrationality only if we fail to be guided, or fail to be guided in the right way, by the (right kind of) rule. The point is that it can be reasonable for us to ignore the rules, or consider them irrelevant, for we can be actively and creatively reasonable: we are not passively confronted with, and just forced to recognise, that which we might consider to be reasonable.

With these points in place, let us consider an instructive recent defence of the rules of logic as super norms articulated by Robert Hanna (2006). Once we have done so, we will be in a position to consider some criticisms of super norms, which often coincide with an approach to the understanding of rationality that is much more akin to the exercise of reasonableness.

---

[77] The opposition between reasonableness and rationality may remind one of terms used in Stephen Toulmin's work, see e.g. Toulmin 2003. It is outside the scope of the thesis to delve into that work in detail.

[78] This is also, incidentally, the form of Glüer and Pagin's 1999 argument against the inherent normativity of linguistic meaning.

A. The 'Intrinsic Normativity' of Logic

Hanna defends what he calls the Kantian constructivism in logical theory thesis (KCLT for short). This is the thesis that 'logic is the result of the constructive operations of an innate cognitive capacity that is necessarily shared by all rational human animals, and governed by categorically normative principles' (Hanna 2006, 68). Hanna calls this innate cognitive capacity 'protologic', and it is this 'single universal' proctologic that is used to construct all logical systems (and has a distinct structure that covers classical and non-classical logic). Strictly speaking, then, it is not the rules of logic (any particular logic) that are super-norms – it is the principles of protologic – but since any particular logic is a necessary result of those principles, one can argue that the status carries over. Certainly, when Hanna speaks of 'the intrinsic normativity of logic' he has in mind logic and not protologic, though he does restrict his claims for there being unconditional categorical imperatives to protologic. In any event, for present purposes what is important is the form of the argument for super norms, not what it attaches to.

Hanna's first task is to connect his picture of logic to psychology. Recall that for Hanna all human beings – 'rational animals', as he refers to us – have an innate cognitive faculty that 'is preconfigured for representing logic, by virtue of its containing a single universal protologic' (Hanna 2006, 70). By rational animals – and thus animals of the kind that can have such a cognitive faculty – Hanna means animals that are 'rule-following, intentional (that is, possessing capacities for object-directed cognition and purposive action), volitional (possessing a capacity for willing), self-evaluating, self-justifying, self-legislating, reasons-giving, reasons-sensitive, and reflectively self-conscious – or for short, "normative-reflective" – animals, whose inner and outer lives alike are sharply constrained by their possession of concepts expressing strict modality' (Hanna 2006, 70). By 'strict modality' Hanna means 'concepts of logical necessity (truth in all logically possible worlds), epistemic necessity (certainty or indubitability), and deontic necessity (unconditional obligation or "the ought")' (Hanna 2006, 70). The virtue of this list is that gives us a clear sense of the kind of picture of human nature / human cognition presupposed by or the result of making an argument for super norms.

To the extent that one presupposes or focuses on 'rational animals', one needs to say something of what one means by rationality. Here, Hanna makes three distinctions, and, in doing so, also refines his own use of the concept of rationality.

The first distinction refers to the difference between the mentalistic and the procedural sense of rationality. The mentalistic sense consists in seeing rationality 'as a complex psychological capacity for logical inference and insight, and also for practical deliberation and decision making', whereas the procedural sense attaches to rationality as a 'complex formal property of a certain class of mechanical, mathematical, computational, or logical processes, namely the property of being (i) well formed and (ii) either provable and recursive…valid…or sound' (Hanna 2006, 71). In the context of this distinction, Hanna says he is interested in the mentalistic sense of rationality. The second distinction is between 'meeting-the-minimal-standards sense' and 'meeting-the-maximal-or-ideal-standards' sense of rationality, where the former 'means either possessing a psychological capacity for rationality or meeting the well-formedness conditions for being a rational procedure of the relevant sort', and the latter means 'either perfectly using a psychological capacity or else perfectly satisfying' the conditions of procedural rationality (Hanna 2006, 72). Here, Hanna says he is interested in the minimal sense. Finally, the third distinction is between the principled sense, the holistic sense and the instrumental sense of rationality, where principled (Kantian in spirit, and also Hanna's interest) means 'the possession of a capacity for generating or recognising necessary truths, a priori beliefs, strictly universal normal rules, nonconsequentialist moral obligations and categorical "ought" claims' (Hanna 2006, 72); holistic (Hegelian in spirit) means 'possession of a capacity for seeing coherence' or reflective equilibrium (Hanna 2006, 72); and instrumental (Humean in spirit) means 'the possession of a capacity for generating or recognising contingent truths, *a posteriori* belies, contextually normative rules, consequentialist obligation, and hypothetical "ought" claims' (Hanna 2006, 73).

The set of distinctions is helpful, and it is useful to see that Hanna's overall interest is in mentalistic, minimal and principled rationality, but there are some ambiguities: for instance, it is not clear what the difference is between so-called maximal standards and the standards of procedural rationality – after all, what could be more demanding, or more maximal, than the process being well-formed, either provable, recursive, and valid (in the sense of truth-preserving) or sound (valid with true premises)? Indeed, the simplest way to put things here is that what Hanna is interested in is how to psychologise those standards of procedural rationality, i.e. how

to show that those standards are part (he wants to say a necessary part – which is why we are looking at him here) of the mind at work.

As an aside, notice the delicate balance between processes and standards in Hanna's choice of mentalistic, minimal and principled. Hanna avoids those readings of rationality that focus either exclusively on standards (e.g. the procedural) or externalise standards so much that it is difficult to see how they could be internalised (e.g. the maximal). Thus, in both the first distinctions, Hanna goes for the mentalistic and minimal, and indeed it is noteworthy that he characterises the minimal/maximal distinction as one that could be read in either a process or a standards sense. In the final distinction, he characterises 'principled' not as a matter of the property of being principled (which would make it identical to procedural) but as a capacity for generating or recognising principles (notice that this can ultimately collapse into the mentalistic if one identifies, as Hanna does, the exercise of capacities for logical inference, practical deliberation, etc as a matter of following the principles). Whenever he introduces processes (psychological capacities), Hanna weaves in the following of standards, and whenever he introduces standards, Hanna does by reference to a capacity to follow them. Hanna, then, is trying to develop a concept of psychology and a concept of logic that walk hand-in-hand. The problem is one we have encountered above: Hanna's approach forces us to think of processes (and psychology generally) as necessarily a matter of following standards. Thus, even if we were to accept Hanna's picture of an understanding of logic that can be psychologised, his picture is such that we would be doing so at the cost of missing all those processes of the mind that are not about following standards. The challenge – which is not open to Hanna given how closely he wants to integrate processes and standards, and because he thinks that there is a realm of super-norms that are unrevisable and innate – is to offer a role for standards, but not a necessary one. This means not only understanding standards in a more modest way (not thinking they have the status of super norms, i.e. that they are necessary / unrevisable), but also understanding the mind (or processes) in a way that makes room both for those processes to be informed by standards as well as not informed by them.[79]

---

[79] One can see here how one can entangle oneself in problems, for if one begins with a definition of rationality as a matter of following standards then it becomes basically impossible to see how else one could make sense of rationality. That is why one needs to begin elsewhere – not with standards – but with the mind at work, e.g. with reasonableness, where normative resources can be used / appealed to, but where they are not the drivers or targets of the mind.

Hanna's device of a protologic is designed to offer a resolution of the debate resulting from the emergence of deviant logics. 'Deviant logic', as Hanna points out, properly refers not to 'extensions of classical or elementary logic' (which preserve, but add to, classical or elementary logic), but to logics that add, delete, or re-define 'classical logical operators, interpretation rules, axioms, or inference rules such that not all the tautologies, theorems, valid inferences, and laws of classical or elementary logic still hold' (Hanna 2006, 75), e.g. relevance logics, three-valued, many-valued, fuzzy logics, and others. The challenge raised by deviant logic, then, is how to think of classical and elementary logic: does one stamp one's foot and say that deviant logic ought not to count as logic at all (and that classical / elementary logic is the one true logic – or of course the other way round, i.e. that deviant logic is the one true logic, but then which one and why), or does one allow for pluralism, recognising different logics (though this, in turn, raises the question: in what sense are all these variations of one thing, 'logic')? Hanna thinks all these options are unsatisfactory, i.e. he does not want to be either a diehard classicist or deviant, or an 'all-embracing' pluralist (Hanna 2006, 78). It should be obvious now where protologic fits in: it offers a fourth solution (or purports to do so), according to which it consists in a 'single set of schematic logical structures, in the form of a coherent repertoire of metalogical principles and logical concepts' that is 'presupposed by every logical system whatsoever' (Hanna 2006, 78). It is an 'unrevisable and a priori' set, which means, at the metaphysical level, that it 'determines what will count as a possible logical system', and at the epistemic level, that 'some knowledge of this set of structures must also be consciously available to thinkers if they are to be able to justify assertions or claims made about any classical or nonclassical logic' – in short, the protologic 'is both constructively and epistemically presupposed by every logical system' (Hanna 2006, 78).

There are a number of questions we can raise at this point. The first is: how does the above solution differ from a pluralist one? Hanna answers this by saying that a 'presupposed logic is not on a par with any of the many logical systems' (Hanna 2006, 79), and that is simply because Hanna defines it as 'a condition of the possibility of there being informative natural language discourse…and reasoning…in the first place' (Hanna 2006, 79). But now recall that the pluralist solution was to say that there are many logics, and none more superior than another, though possibly some are useful for certain purposes than others, or deal with certain kinds of

problems better than others (e.g. fuzzy logic deals better with questions of vagueness – in a way, the justification for fuzzy logic is what it points as needing explanation, which it claims other logics cannot do). To be different to the pluralist proposal, Hanna better have a different take on the hierarchy of the available logics; indeed, he says that 'it is not true that the several different logics are all equally acceptable' (Hanna 2006, 78) – but does this follow in any way from the presence of the protologic? All that Hanna says is that the various logics presuppose the protologic, but does the protologic serve as a standard for ranking the various logics? Notice that the pluralist also has the problem of considering in what sense all these logics are 'logics', but that the pluralist can answer this by saying that they are, roughly, what we tend to refer to as logics – what we, at the moment (as a result of so many years of thinking about logic) think of as logic. Hanna also has the same problem, but he answers this by saying that there is one timeless and non-revisable protologic: he thinks there is 'a core' of logic that all logics cannot reject and yet remain a 'logic' (Hanna 2006, 80).

The point here is that Hanna's proposal can be thought of as a kind of pluralist solution (for he does not tell us in what sense he thinks the various logics are not 'equally acceptable'), but of a very rigid kind. It is rigid because it does not foresee, for instance, the possibility of us thinking of what 'logic' might mean for us in a way unlike any previous way we may have thought of it (this is a possibility because we might create another kind of deviant logic). The open-ended pluralist says that a certain characterisation based on the logics we have is not unrevisable, for it is not necessary that we will continue to think of logic in this way. And, surely, the open-ended pluralist has history on his side: after all, the very question of pluralism did not arise until there were enough deviant logics that it become important to ask what we mean by 'logic' (prior to this, it could be argued we simply meant 'classical' or 'elementary' logic). Another way of putting this is that the open-ended pluralist does not claim (nor feels any need to claim) any higher status (one that points to the criteria being unchangeable) attaching to that which the pluralist offers as a characterisation of what we are currently using as criteria for what to count as a 'logic.' One way to try to salvage something of what Hanna says is to argue that, like Haugeland, Hanna is pointing to the experience we have of treating certain criteria (of what counts as logic) as unrevisable in the course of, in that moment, ranking certain logics – but, as was argued in the case of Haugeland (see Section 7.B.), we over-generalise this

experience if we think that we take in that moment (within a context of systematisation) to be unrevisable must be necessity always be unrevisable.

The difficulties for Hanna here become even more pronounced when we consider what he thinks is contained within protologic. He answers this query in two ways: first, by saying that it is outside the scope of his paper to provide an account of the content of protologic (Hanna 2006, 79), and second, by tentatively offering 'four metalogical principles' which he says 'seem' to him to be 'good candidates for belonging to the protologic' (Hanna 2006, 80). These four candidates, however, turn out to be nothing more than weaker versions of classical or elementary logic: e.g. 'The weak principle of non-contradiction: not every sentence is both true and false' (Hanna 2006, 80). Now the thing about such a principle is that it can be read in two ways: first, as a principle of a protologic that every logic must presuppose; or second, as an acknowledgement of the limits of classical logic. After all, what the principle says – on this second reading – is that that the classical logical principle of non-contradiction will be a good enough guide in many cases, but that there will be cases when it does not apply. Put differently, the second reading helps us see that although this is an important principle, and one we may usefully use as a guide and a standard in most cases, there is no sense in which it is necessary. So why should we accept the first reading? It does not seem to add anything to the second reading and, indeed, it can be misleading for it hides what is effectively contingent application under the banner of necessity.[80]

Let us return to the epistemic claim made above, namely that 'some knowledge of this set of structures must also be consciously available to thinkers if they are to be able to justify assertions or claims made about any classical or nonclassical logic' (Hanna 2006, 78). There is one feature of this claim that should stand out for us, having worked through the above Chapters, namely its prioritisation of the context of justification. In fact, Hanna struggles with how to articulate this. He asserts that 'competent thinkers must be able to (try to) justify the assertions they make' (Hanna 2006, 79) – which includes two different possibilities (being able or trying) – and he also asserts that competent thinkers, whenever they do in fact justify their assertions, will inevitably do so (or will not count as justifying) by invoking

---

[80] We shall ignore here the further argument that Hanna makes for an even more fundamental status for these metalogical principles, as when he says that Chomsky's universal grammar presupposes his, Hanna's, protologic (see Hanna 2006, 81-5) – this simply repeats the same move, and does not add any more plausibility to the first reading.

'conscious logical beliefs about logic' (Hanna 2006, 79), which, on Hanna's definition of logic, means that these must be 'conscious logical beliefs about the protologic' (Hanna 2006, 79). So there are different claims here: 1) that competent thinkers must *be able* to justify their assertions by reference to the metalogical principles of protologic; 2) that competent thinkers must *try* to justify their assertions by reference to the metalogical principles of protologic; and 3) competent thinkers, whenever they justify, must do so by reference to the metalogical principles of protologic.

But do any of these claims work? First of all, it is not necessarily the case that to be a competent thinker I must *be able* to justify my assertions – unless of course one defines a competent thinker as someone who must be able to justify her assertions. The point is that I can think someone is a competent thinker without requiring them to justify their assertions, e.g. I can be perfectly satisfied with their competence as a thinker when they provoke a thought in me, and I need not require them, or think they could, justify their provocation. Second, must I *try* to justify my assertion in order to be a competent thinker – and, if so, competent according to whom? Would this requirement mean that every time I make an assertion, if I also want to be thought of as competent, I must try to justify it? This does not seem plausible: I make assertions all the time, many of which persons might consider competent, without trying to justify them. Indeed, if I were to try to justify whatever I thought I would probably very quickly become unable to think. So the opposite may be more true, i.e. that when thinking, I must be able to avoid being perpetually concerned about the status of my thought. What about the third claim? Is it necessary that when I do justify myself, I call upon the principles of metalogic? What about when I make a claim about a sentence that it is neither true nor false (but, let us say, a guess or a hunch)? In what sense here would the weak principle of noncontradiction serve as a justification for that claim? Hanna would have to say that whatever it is that I use as a justification just happens to be a metalogical principle – but this, surely, is not an informative claim, for Hanna needs to assert, first, what are the principles of protologic, and the point is that as soon as he does so, we can give him an example where no such principle is involved in that actual case of justification. Hanna seems to want to have his cake and eat it: here are, he says, the principles of metalogic, but do not think that just because you come across some instance not capable of being seen as an instantiation of one of the principles I have listed, that there is no relevant

principle of metalogic – you just need to discover the relevant underlying principle. This, once again, is an over-generalisation of the usefulness of appealing to principles of logic (the usefulness, one might say, of such lists is that they are of limited scope: we lose sight of their usefulness when we try to depict them as if they had unlimited reach; our theories, like us, cannot live forever).

Let us now consider, in some more detail, Hanna's specific claims concerning the 'intrinsic normativity of logic.' Hanna does not beat around the bush: he says straightforwardly and confidently that 'Normativity, as such, consists in the fact that there is a set of ideals, standards, guides, recommendations, commands, rules, principles, laws, and so on (hence "norms") that govern human beliefs and intentional actions' (Hanna 2006, 85). Of course, this thesis resists this claim (or assumption?), preferring a much broader account of the normative and, as a result, also a much broader account of the mind (i.e. one that does not depict the mind as necessarily following or being guided by some 'norm' – in Hanna's broad sense of the term). But for Hanna, the mind just is a norm-following thing, and the question that arises is how logic – conceived as we saw above – can feature in the life of the mind. So, when Hanna asks how logic can be normative, what he is asking is how it can be / is followed by persons. And, when he adds that it is intrinsically normative, he is asking how it can be that persons not only can be following / might sometimes follow the principles of logic, but how it is that they must do so. Notice that there is both a pernicious and a non-pernicious (from the perspective of this thesis) version of the question concerning how logic can be normative (without the intrinsic added): the pernicious version implies that the mind must be a norm-following thing, but without the intrinsic allows for the possibility that persons may be following something other than the principles of logic (including when thinking); the non-pernicious version is one that says that sometimes, when we think, we can follow (use as guides) the principles of logic and sometimes not, and when not, this does not mean we are following something else (in other words, the mind is not by necessity a norm/rule-following thing – something that must always be guided by something). It is the second version that we want to adopt in this thesis; Hanna adopts not the first version, but precisely the strongest version possible, i.e. that the mind is necessarily a norm-following thing and that, insofar as it thinks, it is necessarily following the norms of the protologic.

In dealing with the claim of the intrinsic normativity of logic, Hanna takes us on a historical tour of the problem. His way of telling the history of the problem pits the normativists against the descriptivists, i.e. it is assumed that unless one tries to show how the mind is a norm-following thing that follows the principles of logic, one is a descriptivist (and thus someone who can no longer account for the normative dimension). Hanna says that the debate has been between those who claim that prescriptions (moral, logical, etc) are 'determined by actual human interests or natural facts' (Hanna 2006, 91) and those that claim that they are instead 'built into the very idea of practical reason' for they are 'universally binding on rational beings' (Hanna 2006, 91). The dichotomy here is a false one, precisely because both are deterministic: one is deterministic in the sense of persons as blindly subject to certain interests or facts, and the other deterministic in the sense of reason necessarily following certain universal norms. Both can be unravelled by showing, as we have been trying to do, that one cannot posit an underlying or designed order: that any such positing is but a necessarily incomplete specification that treats things to be ordered in a certain way – a specification that can have its uses, theoretically and practically, but does not of itself have any privileged (metaphysical or default) status. The same applies to Hanna: thus, when he says that to say 'that logic is intrinsically categorically normative is simply to say that logic is rationally humanly inescapable, or at least to say that the protologic is rationally humanly inescapable' (Hanna 2006, 95), the equivocation indicated by the 'or' is of great significance, for the very point is that Hanna must commit himself (in a way he does not want to) to setting out what the principles of the protologic are, and then showing how all of human rational thinking cannot escape it; in other words, the very point is that any such specification is going to be limited, and not one that remains an open category of principles that we go on to discover, as if there is a final stage we can reach where we will have finally worked it all out.

In effect, Hanna has no argument save for a plea that these norms (i.e. the principles of protologic) deserve to be treated as presuppositions, as ultimate standards at a higher level, according to which everything proposed either counts as logic or it does not. He is right that these norms (insofar as we can articulate them at this point) are important (for instance, because some specification of them could be a pretty good summary of what we have learnt so far about how we use the term 'logic'); but he is wrong insofar as he thinks that we will never learn more, or that we

could not possibly learn anything that would make us revise the principles of the protologic (the best proof that we can is his very own list, which could not be compiled without having had the benefit of deviant logics).[81]

B. The Rough Ground of Reasonableness

Some of the above arguments against Hanna's position are articulated very well in a number of papers that resist this elevation of logic (or protologic, in Hanna's terms) to the status of super norms. Indeed, the term 'super norm', as used here, comes from Peter Railton's excellent essay, 'A Priori Rules: Wittgenstein on the Normativity of Logic' (2000). Other sources friendly to this position include Jane Heal's (2007) argument for a return to 'the rough ground' of the exercise of reason, and Stephen Turner's characterisation of Davidson on rationality (in Turner 2010, chapter 6). Let us look briefly at these sources, as a way of rounding off this Chapter, and indeed this Part of the thesis.

Railton confesses from the beginning his uneasiness with the 'a priori', calling on 'the cold shadow of history' and the promise of change in 'the future course of experience' (Railton 2000, 170). Of course, saying that 'anything goes' is not a tempting option either. The trick is to find a balance between indifference and absence of discipline on the one hand, and stubbornness and rigidity, or the bearing of a mental straightjacket, on the other hand. At stake, Railton says, is the possibility of 'adaptive intelligence hard at work', or 'adaptiveness' in short (Railton 2000, 171). Railton's paper treats a number of problems related to the *a priori*, but for present purposes the interest in his paper is what he says about *a priori* standards, or so-called 'super norms.'

One of the impressive features of Railton's paper is his sensitivity to the variety of attitudes that may inform our relations to standards. As he points out, there is a seriousness and humility – one might add courage too, and honesty about oneself – in Kant that we miss if we read into him only rigidness: there is a sense in which

---

[81] There are other aspects of Hanna's paper we have not considered, e.g. the argument against the claim that his picture cannot account for error (Hanna 2006, 98-100) – we have ignored it here because it assumes that error consists in the violation of standards, which we have already dealt with in Chapter 6 above. One comment can be added: Hanna laments the making of errors, and, indeed, if one defines error as violation of the rules of the status of protologic, then there is much to lament; there is less to lament, and more to celebrate, when we realise that what counts as the making of an error from Hanna's perspective may sometimes be a matter of finding / creating an alternative way to do things, of new cases to consider, new problems – the point is that violating the rules of the protologic is not necessarily something bad, as Hanna assumes it is.

rules, as external constraints that we autonomously acknowledge, can help us learn to be independent (not be swayed by the crowd), avoid temptation, discipline gusts of desire, and also to look at ourselves without immediately succumbing to that which we might otherwise 'naturally' (out of self-interest, say) feel like doing and without the kind of self-justification (a kind of moral hypocrisy) that our desires can talk us into. The question is: is there something more than this, and / or does there need to be? Do these rules that we may rely on for the above – rules that we may (come to) care about deeply, especially where these are moral rules – have a status that renders them immune from doubt, safe from challenge, perpetually adequate, necessarily unrevisable (etc)? What are the attitudes that we miss (the importance of) when we think that all we need is the discipline to commit to certain rules?

Railton's paper quickly comes to the issue of the rules of logic, especially classical propositional logic, and to the argument – which we encountered above with Hanna – that these rules establish 'the limits of real possibility and real thought' (Railton 2000, 175). Just as quickly, the focus turns on what sense may be given to these rules as being both *a priori* and as guides to conduct (i.e. more strongly: does showing how they can guide in fact reveal that they cannot have the status of being unrevisable?). To help us, Railton brings us down to earth (literally) with a few images: the term 'rule', he reminds us, comes from the term 'regula', meaning ruler or straight edge (and etymologies are offered also for norm, standard and correctness: see Railton 2000, 179). Railton's paints the following picture:

> The mason or carpenter initially confronts rough workpieces – lumber or stone – to be cut to size for their place in the construction to be made. Prior to making a cut, the workman places the *norma* [the square] or *regula* [the ruler] against a workpiece as a guide, and scribes a line for his cut to follow. He then cuts to the line, and afterwards reapplies the *norma* or regula to test the cut for squareness or straightness. If there are gaps between tool and workpiece, the cut will be 'brought into line' or 'corrected' (*corrigere*) using a saw, chisel, raps, or plan, until the fit is tight. Only in Chaplin comedies will the *norma* or *regula* itself be 'corrected' to match a crooked cut by filing away at the tool until it fits. (Railton 2000, 180)

Further, insofar as the builders (masons, carpenters, and others) are working together to build something, they need to co-ordinate their efforts, i.e. they need to (presumably this is not all they need to do, but part of it) commit to relying on the same tools, the same standards of measurement, etc. If they do not, it is unlikely they

will be able to arrive at completing a building that is stable enough to live in. In a sense, one could argue that the builders need to commit *a priori* to the tools they use – if only because there is no sense in which conformity with the rule (after the fact) or violation of the rule (also after the fact) either confirms or disconfirms the rule (it just serves to evaluate the performance; Railton 2000, 180). Any builder that, halfway through building the house, began to use a different tool, or a broken one say, would be roundly criticised, and probably fired (perhaps even the other builders, in order to emphasise their condemnation of the builder, would say he was no longer 'really building').

Now if one stopped here one could see how one could come to think that there are such things as tools (or rules, standards, norms etc) that have such an important standing that they deserve being called *a priori* (if anything does).[82] But Railton goes further. The point he makes (by reference to yet another tool, the contour gage) is that the tool itself does not determine when it ought to be used. The *norma* and the *regula*, as Railton puts it, 'lack "inherent power" to guide' (Railton 2000, 182). There are, in fact, three points one might find in this observation: first, the tool does not itself determine when it ought to be used (when it is relevant); second, it does not itself determine how it ought to be used (how it is to be applied); and third, the tool itself does not secure its own appeal (its motivational force, one might say). Seen from this perspective, does it make sense to call such tools *a priori*?

To Railton's example we may add others; for instance, imagine that the builder who, half-way during building the house begins to use a broken tool, is confronted by his fellow builders, and imagine that he explains to them that when the tool broke (let us imagine it was a square), and he welded together two of the sides, he found that it helped him, in a more effective fashion, to calculate angles exactly half-way between two-sides of a square. It is, of course, possible that he would be fired, but it is also possible that he would be introduced to the architect, and rewarded with a healthy bonus – especially if the device helps saves time. Notice that it is also not necessary that for the builder to achieve a bonus (because he performed his work quickly) he would need to tell someone about his discovery – perhaps he got the job

---

[82] There might be room for argument that this is a weaker form of *a priori* than the one attributed to the rules of logic vis-à-vis thinking, for here we may decide not to build a house, or a house of that kind (that needs such-and-such tools), whereas it is difficult to say that we can decide not to think. But if even this weaker form of the *a priori* (if it is weaker) does not stand up to critique, then the stronger one is likely to come off worse.

done quicker all by himself, and in secrecy. No doubt, in such a scenario, there is a danger that he would get some calculation wrong (for he would need to halve the angles each time, etc), but it also possible he would not. The question here would be: would thinking of the tools as *a priori* mean that we exclude from possibility the case of the inventive builder, and thus also exclude the possibility of finding an alternative – perhaps more efficient – means to the same path?

To this example we could add another: imagine that half-way during building the house another builder discovers, quite by accident, that if the roof were to be sloping, then the snow would not remain on it and risk cracking it (imagine that he leaves bits of the roof at an angle and sees this happening). Imagine that the builder would then go to the building supervisor and tell him his idea: 'we could make it slope, sir!' Now imagine that this is in fact impossible to achieve with the tools the builders have. How would the building supervisor react? He could say: 'look, that's a great idea, but it's impossible, and we don't have much time, so get on with your work'. But he could also say: 'ah, what a great idea, let us try to think how we can do that, what different tool we might need to invent for the purpose'. Now the point is that insofar as we raise the importance of the tools to some *a priori* standard – not only unrevisable, but so high up that they determine what can count as a tool – then the second example is also excluded as a possibility.

As Railton points out, it follows from cases such as these that a 'genuine super-norma would or could not fail to anticipate all needs and situations, and all innovations in building materials or techniques' (Railton 2000, 184). Are there such super norms? Are there norms to which we ought to stick to religiously, 'come what may' (Railton 2000, 185)? Of course, answering 'no' to these questions does not mean that we cannot have the experience in certain situations that something ought to be treated as if it were a super norm – perhaps we think it so incredibly important to use it as a guide or standard of evaluation that we are willing to put all our eggs in one basket in that situation. But no such experience itself has the power to turn a norm into a super norm, or to be so generalisable as to apply in general to some norm. Railton says as much when he says that 'we need to be able to regulate our practices by *normae* that fit various purposes and can be used as standards for our often actual imperfect performance', as long as we do not take this need so far as to assert that it requires 'us to think it is unconditional in application or inherently thought-guiding' (Railton 2000, 194). There are three small points to add to this: first, we need not

think that just because we have found some norm to fit some purpose well, that it is necessarily tied to that purpose (and that therefore no other alternative means might ever be found); second, norms (in general, not just some specific set of norms) are not the only way in which we can discipline ourselves (though they may be a particular efficient method for doing so); and third, if we appeal to purposes, it is important not to think these are fixed (that there are some set purposes such that we are confined to finding ever-better means for achieving them or pursuing them).

Similar themes play themselves out in Jane Heal's paper (2007). The paper begins with a sketch of 'perfect rationality', going on to show why it such a concept of rationality cannot 'get a grip on us, i.e. set[] a standard for actual human beings' (Heal 2007, 404), and rounds off with guidelines for a 'more modest and applicable' concept of rationality (Heal 2007, 404). Like Railton, Heal quickly comes upon the topic of the rules of classical propositional logic (though she also mentions decision theory and game theory). Heal asks: 'Can we make sense of there being "what perfect rationality demands of us"?' (Heal 2007, 409). Heal argues that we cannot because 'there is no given, context independent, set of beliefs and desires for the demands of rationality to bite on' (Heal 2007, 412). The idea here is 1) that the demands of perfect rationality require there to be a 'set of a person's beliefs and desires at a time', i.e. some 'determinate starting point in us from which the demands are generated' (Heal 2007, 409), but 2) that this requirement cannot be fulfilled because a 'person's concepts…are not given once and for all. They are open to questioning, development and replacement' (Heal 2007, 410). Heal here articulates this second point by suggesting that our concepts always depend on presuppositions that are not articulated or explicit (and thus cannot be 'grist for the formal mill' (Heal 2007, 412). Heal's observations are important, but the problem with this way of emphasising 'questioning, development and replacement' is that it leaves the door open for an argument for implicit rule- or norm-following (of following presuppositions that we have not yet made explicit), which is a direction that (as the next Part of the thesis argues) we should avoid.

Having dismissed perfect rationality on the basis that it cannot get a grip on us, Heal looks around for an alternative – something 'more modest and applicable.' As she does so, she considers what kinds of considerations (qualities of persons) should guide us in this search. Here, she says:

Human beings are ingenuous and inventive animals. Given time and resources (and a cultural climate not wholly against innovation) humans will develop and elaborate their techniques and practices. We did it with ways of knapping flints, shaping fishhooks, building houses, making garments and smelting metals, where the archaeological record shows the many lines of experiment and development which have been worked through. The overwhelming probability, then, is that our linguistic practices, including the practices in the area we label 'debate', 'discussion', 'argument', 'persuasion', and the like, have undergone similar elaboration. There will have been attempts to classify, systematise, improve, elaborate, carrying on over the centuries. (Heal 2007, 414)

Heal also speaks of variety, pointing to the fact that 'people may be struck by different aspects of "good reasoning", depending on what kinds of debates they first bring into focus' (Heal 2007, 415). Her sensitivity to historical contingency (both in terms of development and variety) is visible everywhere on the pages of her paper.[83] Informed by this sensitivity, she offers the following alternative (to perfect rationality): 'conversability'. A 'conversable being', as she puts it, 'is a being with whom one can (at least sometimes) enter into dialogue' (Heal 2007, 417). This is an important idea, and Heal comes close to recognising that something special may be on the horizon here if one examines the dynamics of interaction from a second-person perspective. But she opts, in the end, for something else, focusing matters on 'reaching agreement or understanding' (Heal 2007, 417), on justifying one's responses (coming up 'with relevant supporting moves for her claims or proposals'), and on 'being familiar with insights into good methods and appropriate standards of reflecting and discussing', with the aim of developing an 'ability to converse impressively by the standards of [one's] society' (Heal 2007, 418).

Certainly, being able to justify oneself, and to care about doing so, and to be familiar with standards in the ways that Heal suggests – including even occasionally having the aim of meriting aplomb from one's peers – are all important qualities, but they are hardly qualities that echo Heal's earlier notion of 'ingenious and inventive animals.' We need other qualities here, such as being willing to entertain doubt (or having and sustaining the state of uncertainty), which may involve questioning (rather than seeking to emulate) the 'standards of one's society' and not necessarily aiming for 'agreement and understanding' with others in the course of dialogue with them.

---

[83] To this sensitivity we ought to add sensitivity to the variety and development of ways of thinking about history, telling the past, etc – for a recent *tour de force* of such variety and development see Burrows 2007.

But we will need to suspend discussions of these qualities, for we shall be returning to them later in the thesis.

Stephen Turner's (2010) reading of Donald Davidson adds other elements to this momentum of alternatives to thinking about rationality as exhaustively defined by a set of super norms. Turner replaces the idea of rationality with intelligibility (or, put differently, says that intelligibility is all there is to rationality), and further elaborates on the idea of intelligibility by cashing out the notion of 'following the thought of another.' Turner's discussion of Davidson tries to rescue him from those who would wish to read him as defending some concept of rationality distinct from intelligibility – which tends to be accompanied (for instance in McDowell and Brandom) by some metaphysical status being given to certain norms. By contrast, on Turner's reading, Davidson 'lacks the Brandom-McDowell imagery of constraint' (Turner 2010, 163). What this means, for instance, with regard to error is that it is not thought of in terms of 'deviation from rules…there is no constraint here, no disciplining by group reactions', but rather in terms of 'the problem of making oneself understood to other individuals and of understanding other individuals' (Turner 2010, 161). On Turner's understanding, Davidson's arguments are about the uses and limits of the intelligible and that is it – there is no appeal to anything allegedly more 'fundamental' or 'inescapable'. What matters for Turner's Davidson is that 'we want to be able to follow others, to follow their reasoning', which we do by 'a hypothesis-testing epistemic process, in which we employ what we know about ourselves and our beliefs to construct accounts of others' beliefs that make sense of their behaviour' (Turner 2010, 162; this includes the exercise of interpretative charity, see Turner 2010, 157). These hypotheses may involve theories of rationality, but these are not indispensable in the way that the 'normativist' (as Turner calls the position) would like to read it, namely as 'necessary, and necessary in the manner of synthetic a priori truth' (Turner 2010, 166), but rather 'indispensable for our various theoretical and even practical purposes' (Turner 2010, 167). Here is perhaps Turner's clearest statement of how he understands Davidson's concept of rationality:

> Recognising something as rational is a matter of being able to follow someone's thought – to simulate his or her thinking well enough that this individual's differences can be allowed for as normal or explained, and thus made intelligible as error. The normative element is not rigidly fixed, unarguable, or even free from conflict between the kinds of inference that we

can follow but which also lead to conflicts. This is not the kind of rationality that provides the kind of constraint and ultimate justificatory ground that is the concern of Brandom or McDowell. The only constraints are interpersonal: we are constrained in our understanding by the limits of what we can follow and we are constrained in communicating by the limits of what others can follow, and constrained in what counts as thought by the requirement that for something to be recognised as thought, it must be the kind of thing the recogniser can follow. (Turner 2010, 167)

Some might think that Turner goes too far in identifying rationality with intelligibility. Further, to the extent that Turner wants to rid the notion of any connection with the 'imagery of constraint', as he puts it, this does seem to throw the baby out with the bathwater. For, as we have seen, there is nothing pernicious (or 'normativist', in Turner's language) with the idea that we can, sometimes, guide ourselves by normative resources and use them to evaluate others (including tying certain normative resources to our use of the term 'rationality'). One need not argue that the only normative resources that exist are the constraints of intelligibility, for there can be normative resources that we can associate with and that can inform (but not determine) our evaluations of something as rational (after all, there does some to be some difference in the experience of evaluating something as rational and finding something intelligible). The point is that we need not collapse rationality into intelligibility in order to pull the metaphysical sting out of rationality. Similarly, we need not banish the concept of justification just because it has often been used in the sense of an 'ultimate justificatory ground' – we just need to see the limits of what a context of justification can explain.

Further, it is not clear that intelligibility, as Turner understands it (based on the idea of following the thought of another, which he adds 'is perhaps best understood in terms of the idea of simulation in cognitive science': Turner 2010, 168), can do the work Turner wants it do in social contexts. Turner speaks throughout the book – and this is one of its most admirable features – of the need to resist a sense of social that collapses into the 'necessarily shared' or 'necessarily collective', or 'the idea of shared frameworks' (Turner 2010, 168 – see especially chapter 5), and he urges us to embrace the notion of interaction, of 'the interactional sense of the social' (Turner 2010, 168). This is an excellent suggestion, which we will be developing later in this thesis, but what is noteworthy here is that this notion of interaction is missing from the concept of intelligibility, and indeed from the simulationist tradition in cognitive

science. Turner rightly resists the kind of third-person stance he sees emulated in what he calls 'normativism', but he seems to replace it only with the projection of one's own acquired knowledge (a first-person stance) – a view of the social world where we only interact in the sense that we see each other as performers in schemes of intelligibility (even if these are said to be free from being necessarily shared).[84] Turner is surely right to resist attributing metaphysical status to theories of rationality (Turner 2010, 167), and he helps us by pointing to the context of interaction, but this context is one that needs to be looked at in ways that go beyond the first-person / third-person divide (and thus beyond the opposition between shared frameworks and simulation).

Railton's, Heal's and Turner's observations can help us to paint the following picture: at bottom is the rough ground of reasonableness, an indeterminate ability that is active and creative, i.e. we can create what we might go on to treat as a standard of reasonableness (this does not mean we can afford to do this all the time, and that anyone can afford to do so whenever they wish). But at any one time, we can (and this is also an active and creative thing) treat some standard as being unrevisable. We might do so, for instance, in a context of justification where we need to justify a claim as to why something is or is not rational. In other words, rationality is built on top of reasonableness: the clean, pristine world of the rational is best understood as a contingent (active and creative, and not default) attitude, exercised where there is some benefit to be gained from treating things as being determined in some way, and tightly controlling the environment in that fashion. The default is indeterminate: it is the rough ground of reasonableness, where we explore how things might be reasonable, and where we can appreciate something to be reasonable in a way we had not previously ever appreciated. The switch makes all the difference, for we allow for the occasional guidance and evaluation (of ourselves and others) by what we can treat as unchangeable rules (e.g. the rules of logic), but do not make those rules what we must necessarily (either by regular default, or in fact unavoidably) follow. This is not a thesis about reason, but if it were what it would need to explore is precisely this priority of the reasonable over the rational, and the circumstances in which we build determinacy (the rational) on top of indeterminacy (the reasonable).

---

[84] Note that here Turner is referring to 'stances' in a different way than this thesis refers to the second-person perspective. Turner is talking about the stance an agent takes in relating with another person. The second-person perspective is a way of seeing what is importantly part of social life: it is a device used by a theorist; not a stance of a social agent.

We have come to the end of this Part of the thesis. Although it has taken us a lot of time, this was time that needed to be spent, for one of the most difficult arguments to shake off when tackling the governance view is the idea that determinants of appropriateness and inappropriateness are necessary. It has been argued above that such determinants are not necessary, whether we are thinking about the very idea of appropriateness or the possibility of error. Similarly, arguments for the necessity of determinants in constitutive form, whether applied to the alleged nature of games or of language, also do not work. Finally, there are no such things as super norms. None of this means, as has been argued at length, that we need to dismiss the notion of normative resources (the terminology is important – we need to shift from thinking of conventions, norms and rules as determinants, and see them as contingently useful resources). We just need to see that they are of contingent usefulness, while acknowledging that, in certain contexts and for certain purposes, we can experience them as necessary.

The rescuing of the contingency of normative resources, and the distancing from the necessity of determinants of appropriateness and inappropriateness, is a crucial part of making room for the second-person perspective, and thereby also broadening out the canvas of the social and our abilities to participate in it. It is hoped that by proceeding slowly, through four different versions that the first element (necessity) of the governance view can take, we have given some reasons to doubt this initial, critical plank of the governance view, and made plausible the need for a point of view that takes seriously the tentativeness, that mix of the predictable and unpredictable, the balance regular and the irregular, that lies at the heart and root of the social and our ability to participate in it.

Part III. Governed Minds:
The Second Element of the Governance View

Although, in this third Part, we are still focused on articulating and critiquing the governance view, we need to change how we are approaching it. This Part does not, as the preceding Part did, tackle the governance view from the perspective of the necessity for determinants of appropriateness and inappropriateness. Instead, the focus is on how characterising the mind as a matter of being governed by those determinants has a deleterious effect on our understanding of the mind, i.e. how it narrows our picture of the mind and how it works.[85] Insofar as we speak of narrowing, we will of course also be pointing to what is being missed, and that will mean that we will foreshadow, to some extent, aspects of the positive picture we will be offering in Parts V and VI.

In this Part, we will be examining various manifestations of the idea – given the assumption that determinants of appropriateness and inappropriateness are necessary – that the mind is perpetually, and by default, guided in some way (mainly, but not only, by conventions, norms and rules). Not all theorists speak in terms of 'guidance', and indeed there is no neutral term here.[86] Perhaps common to all these positions is the idea that the mind is somehow oriented (and learns to be oriented so as to be able) to recognise something it ought to recognise, where what it 'ought to recognise' is said to be determined by some determinants of appropriateness and inappropriateness.

We will be exploring five manifestations of this second element of the governance view in five Chapters. Of course, some of the issues that arise will overlap, but the examples and arguments covered in each Chapter are still distinct enough. Chapter 10 examines the rule-following literature, though not in the sense of an exhaustive history, but rather by way of picking out leading examples and ways of dealing with them, and analysing them critically. Chapter 11 looks at the notion of

---

[85] 'Mind' is here being used neutrally as far as possible, and thus not implying any dualism between mind and body.

[86] The term 'guidance' itself is harmless, as long as it is not said to be a default mode of relating with others and the environment. As we will see later, some of the explanatory burden in theories that appeal to the way the mind is guided is moved to an account of the potential for a critical reactive attitude, i.e. it is said that the mind is guided by a certain normative resource, even without realising it, when it would (if the need arose) use it as a ground for criticising (itself or others). When reference is made to guidance, the primary meaning is to guidance without positing a potential critical reactive attitude, but in some theories this separation is artificial.

computation, including in theories of the acquisition and exercise of linguistic ability. Chapter 12 turns to a view of the mind as necessarily conceptual, where conceptual itself tends to be understood (on the Kantian model) as rule-governed (the focus here is on the work of John McDowell). Chapter 13 considers the work of Erik Rietveld who has recently promoted a research program based on the possibility of action being adequate, yet unreflective. Finally, Chapter 14 turns to those accounts of tacit knowledge (principally a recent one offered by Harry Collins) that also present the mind as oriented by determinants if appropriateness and inappropriateness, though ones it does not know and possibly could never make explicit.

In all these theories, the focus is on arguments designed to show how the mind is actually guided, or really guided, or really oriented normatively – in the sense in which it has learnt to be disposed to, attracted to, sensitive to, etc, determinants of appropriateness and inappropriateness. We shall try to make these views as plausible as we can, and also consider what of them we can accept, while nevertheless looking very carefully at how they narrow our understanding of the mind.

To clarify a possible misunderstanding from the beginning, let us reiterate that the claim here is not entirely dismissive of those theories that account for the mind can *sometimes* be guided – the idea, instead, is to resist over-generalising this to such an extent that it is claimed that the mind is perpetually and by default guided by a certain set of determinants of appropriateness and inappropriateness; that it is, in that sense, governed by those determinants. In moving away from this over-generalisation, this Part asks us to consider the possibility that we may move through the world, experience it and engage with it, without having something before us that we always and already recognise, and are pre-disposed to recognising, in accordance with some criteria that we somehow always and already possess. As noted above, we shall see glimpses of this positive claim in this Part, but a proper elaboration of it must wait till Parts V and VI of this thesis.

The mind, from the second-person perspective, cannot be governed in the way that the governance view depicts it as governed. Insofar as the governance view depicts the mind in that way, it risks not being able to see, let alone describe, the abilities that persons need to participate in social life, including in its normative dimension (for instance, having non-convention, non-norm and non-rule based normative experiences in the course of interacting with others). It is, then, vital that we are able to show the inadequacies of an approach to the mind, or cognition, that

sees it as being necessarily governed by determinants of appropriateness and inappropriateness.

Chapter 10. Varieties of Rule-Following

This Chapter introduces the variety of ways in which theorists have dealt with the issues thrown up by accounting for the mind as a rule-following thing. It is a long Chapter, but this is because it tries to give a detailed enough flavour to the variety of ways in which the challenge of rule-following has been met. We will begin with a note about the historical background to the debate, and then proceed to examine various attempts at resolving the issue, focusing on those that appeal to the notions of inclinations, dispositions, abilities, and blind / implicit or other unconscious forms of rule-following.

A. The Debate and its History

The historical background to the debate is usefully articulated and reviewed by Paul Boghossian's influential 1989 paper, 'The Rule-Following Considerations'. As Boghossian points out, enthusiasm for the topic emerged in the wake of Kripke's *Wittgenstein on Rules and Private Language* (1982). In that work, Kripke engaged in great detail with passages on rules and rule-following in Wittgenstein's *Philosophical Investigations* (1953) and *Remarks on the Foundations of Mathematics* (1983). Kripke famously attributes a 'sceptical conclusion' to Wittgenstein, which he articulates as follows: 'There is no fact about me that distinguishes between my meaning a definite function by "+"…and my meaning nothing at all' (Kripke 1982, 21, quoted in Boghossian 1989, 508). Kripke finds his way to this conclusion by eliminating numerous possibilities for showing otherwise, e.g. by relying on 'facts about how the speaker has actually used the expression, facts about how he is disposed to use it, and facts about his qualitative mental history' (Boghossian 1989, 508). The most controversial of these eliminations has turned out to be the argument against the dispositionalist account.

According to Boghossian (1989, 508), Kripke had two arguments against the dispositionalist: first, Kripke argues that whereas the meaning of words is infinite, dispositions are finite, and so dispositions cannot account for meaning; second, he argues that dispositions cannot account for the normative dimension of meaning, for dispositions – as predictions about what I will do – will not do as accounts of what I ought to do and this is precisely what an understanding of meaning requires (hence the slogan that 'meaning is normative'). Discussion of the arguments against the

dispositionalist have tended to focus on how exactly to articulate the second argument, namely how exactly to make sense of the normative dimension of meaning. Ironically, as we shall see, some of these attempts throw us back into what was arguably the (or one of the) target(s) of Wittgenstein's remarks: that rules and rule-following cannot account for meaning.

Boghossian's paper focuses precisely on proposals to understand the second argument against the dispositionalist, or, more positively, on proposals for accounting for the normative dimension of meaning. Ideally, for many, such an account would also show us a way of avoiding the sceptical conclusion.

Some of the issues raised by Kripke are ones that would return us to issues we have discussed above. For, on certain understandings of the 'sceptical challenge' (as articulated by Kripke on behalf of Wittgenstein), the task is to show how 'having a meaning is essentially a matter of possessing a correctness condition' (Boghossian 1989, 508), where that correctness condition determines what is a correct and incorrect use of a word. That Boghossian thinks that what is at issue in the normativity of meaning thesis is precisely whether meaning can be determined by criteria of correct use is visible throughout his paper, e.g. he says that 'Kripke's insight was to realise [the following]…condition of adequacy on theories of the *determination* of meaning: any proposed candidate for the property in virtue of which an expression has meaning, must be such as to ground the "normativity" of meaning – it ought to be possible to read off from any alleged meaning constituting property of a word, what is the correct use of that word' (Boghossian 1989, 513; emphasis added). He adds that it is because there can be such a thing (as the determination of meaning by criteria) that the dispositionalist can fail to account for meaning, 'for, it would seem, one cannot read off a disposition to use a word in a certain way what is the correct use of that word, for to be disposed to use a word in a certain way implies at most that one will, not that one should' (Boghossian 1989, 513).

Later, we will see resistance to this way of reading dispositions (Boghossian himself asserts that he is merely trying to state the objection, not defend it here), but it is relevant to observe how the debate is set up, i.e. between the right kind of determination (normative) and the wrong kind (predictive). This is a problematic way of setting up the debate from the perspective of this thesis, because what we have been arguing is that it is not necessary to think that something (in this case, meaning) must be determined for that something to have a normative dimension. This does not

mean that the anti-deterministic position is a sceptical one – one must avoid trying to silence the sceptic (the sceptic's voice is an important one); one can acknowledge the limited sense in which the sceptic's voice is important, but yet not fall into scepticism.

The problem tackled in this Part of the thesis, however, is not with the issue of whether correctness can be (and ought to be thought of as needing to be) determined by criteria, but what effect thinking that it can, and trying to show that it can (by 'rescuing' rule-following), has on our picture of the mind. It is a little unclear in Boghossian's paper as to whether he thinks that the problems thrown up by rule-following are identical to the problems concerning the possibility of correctness conditions (by 'correctness conditions' Boghossian means precisely the possibility of – and indeed necessity for – there being conditions that determine what is correct). He says, for instance, that 'exploring the possibility of correctness is tantamount to exploring the possibility of rule-following' (Boghossian 1989, 516), but then later he says that 'the ordinary concept of following a rule…presupposes the idea of a correctness condition, not one that can, in full generality, help explain it' (Boghossian 1989, 517).

It is true that the motivation for tackling the rule-following problems may be a motivation to show that there are correctness conditions, but it does not follow that the two sets of problems are identical. Further, the concept of following a rule can be divorced from the question of whether correctness requires conditions to determine it and whether there are such conditions. In other words, we can maintain that sometimes we do follow rules, and yet not agree that correctness is the sort of thing that requires determinants and that there are determinants. Certainly, when we argue that we always and necessarily follow rules (e.g. whenever we mean something by a word) then there may indeed be an overlap between the reading of correctness and the rule-following considerations. The point for our purposes is this: we have to keep in the back of our minds the possibility that an account of following a rule can be motivated by the aim of showing that correctness requires conditions to determine it, all the while nevertheless keeping our focus not on that motivation (and thus on whether the aim has been satisfied) but rather on the plausibility (including scope) of the account of how we allegedly do follow rules. In light of that, perhaps what we should say is that Boghossian has things the wrong way round: it is not that the rule-following challenge presupposes correctness conditions, but that those who are

committed to there being correctness conditions, and to correctness requiring them, presuppose that we are (always, rather than merely sometimes) rule-followers.

B. Non-Intentional Rule-Acceptance

Boghossian has helped us glimpse some of the historical background to the issues,[87] but what is his own position? The 1989 paper does not help us – in terms of our focus here on rule-following – for, as mentioned above, it remains at the level of whether there are correctness conditions. We need to look, instead, at some of his later work. But before we do so, let us very quickly see what he does say about correctness.

Boghossian asserts that it is 'incontestable' that 'The fact that "horse" means *horse* implies that "horse" is correctly applied to all and only horses' (Boghossian 1989, 530; original emphasis). The use of quotation marks and italics should already suggest to us that there is something worth investigating here. Boghossian helps himself here to a distinction between 'horse' and *horse* (this looks like a distinction because of the typographical differences), while at the same time saying that they are one and the same thing. The concept of *horse* suggests that there is something that of itself determines whether it ought to be called "horse." And this is indeed how Boghossian sets up the cases: imagine, he says, that on a dark night we look at an object and because of the circumstances call it a "horse", but it in fact is a *cow* (it just looks like a *horse* in these circumstances) (Boghossian 1989, 531). These cases make it look as if it is obvious that one can use "horse" correctly (i.e. when it really is a horse), and incorrectly (when it is not), but also that the dispositionalist (at least a minimal kind of dispositionalist) cannot account for this difference, for one can have a disposition to call "horses" cows-that-look-like-horses-in-dark-nights. On this way of setting things up, the dispositionalist cannot account for error, for whatever one is disposed to do with a word is, for the dispositionalist, the meaning of the word.

There are a number of assumptions and case-descriptions at work here. The basic problem is with how the above conceives of criteria. Basically, the cases are set up in such a way that all the pressure is on us, as followers of the criteria, and no pressure is placed (indeed, ever conceived as possibly placed) on the criteria themselves. Thus, the cases make us forget that the criteria themselves can be questioned and that following them can be the cause of a mistake. To see this,

---

[87] For a useful historically situated discussion, especially in terms of Wittgenstein's remarks, see Schulte 2007.

imagine a situation in which one of us notices that some things we have been calling "horses" have spots under their belly, and some do not (it is an interesting question *when* a person might 'notice' this, but we shall leave this aside for the moment). The criteria we have used to identify horses (assuming for a moment that this is the only way we do identify horses, which needs to be resisted, as shall be argued later) have previously not distinguished between horses with spots and those without spots. As soon as we see this, we also see that the criteria are finite – indeed, they need to be finite in order to help us identify some phenomena rather than another. Their being finite means that they set up certain differences and ignore others – in a way, their assistance in identification means precisely that they organise differences in a certain way (they make some differences relevant, and some not). Now the point is that when faced with the discovery that some beings we have been calling horses have spots and some do not, we need to decide whether to continue ignoring the difference or to create two different kinds ("horse" and "shorse") or perhaps a sub-kind ("horse with spots" and call that a "shorse"). How we proceed here is not determined by anything in advance – it may be informed, for instance, by the purpose of our classification or simply by how we use horses (imagine that we try to test whether having spots makes horses stronger or weaker for some purpose).

Boghossian's cases, then, neglect the possibility of revising the criteria, for they set things up in such a way where the only relevant error is one of us violating the criteria – the criteria are assumed to be infallible and to some extent infinite (they apply automatically to whatever really is a *horse*). Such unrevisability in the criteria can only be so if we also assume unrevisability in what can matter to us or that we have learnt all there is to learn about the world such that we have decided, for all time, what should count as a relevant difference.

Boghossian is right, however, in expressing a worry about a situation in which we would never have access to anything external to our own performance, such that our performance (our 'say-so') would determine what is correct and what is not. He is right that to some extent we need, and it is helpful to have, criteria, but where he goes astray is in giving us no opportunity to see (because of the way the cases are described) that any criteria are not applicable by default, as if they will never need revision, or as if they were always relevant unless we are given reason to think otherwise. The point is that we are not passive vis-à-vis relevance. It is precisely because it is up to us what may or may not be relevant that normative resources are

contingently useful, and also finite – we may, on occasion, think them relevant and apply them, but we may also think them irrelevant, and go on to create relevance. In describing cases, then, we always need to leave room for the possibility of not only questioning the criteria (or normative resources in general), but also of creating relevance in a way that makes any previous criteria redundant. The cases cannot be described in such a way that they determine what is relevant, as if everything that we encounter is destined to be treated by us as either in conformity or in violation of those resources (as if, precisely, we were always – by default – guided by those resources).

In a way, even this brief discussion shows that if we think correctness requires determinants, and thus we forget how we can actively create what might be relevant and what might matter for us, then we might also see how we cannot be exclusively or even dominantly rule-followers. In this Part, however – to repeat what was said above – we are trying a different strategy, namely one that shows how implausible are the consequences of our understanding of the mind learning and working when we picture the mind as only or even mainly a rule-following thing. Let us return, then, to more recent work by Boghossian where he does tackle issues more relevant to problems to do with rule-following.

In his most recent paper on the topic, 'Epistemic Rules' (2008a), Boghossian tackles not meaning, but belief (or more accurately still, what he refers to as 'rational belief'). In doing so, he calls on a category of rules he calls 'epistemic rules', which he says 'tell us in some general way what it would be most rational to believe under various epistemic circumstances' (Boghossian 2008a, 472). An example of one such rule is 'If it visually seems to you that p, then you are *prima facie* rationally permitted to believe that p', but he also includes more traditionally referred to rules (such as *modus ponens*; Boghossian 2008a, 472). We will not here evaluate Boghossian's claims as to the status of these rules (e.g. he says that they 'constitute what [we] may call our epistemic system': Boghossian 2008a, 473).[88] Instead, what is of interest here is how he establishes 'what it is to follow a rule' (which he says he wants to do 'regardless of how exactly a rule is construed': Boghossian 2008a, 474), and what effect this has on our understanding of the mind.

---

[88] To do so would return us to the issues discussed in Part II, especially Chapter 9.

Boghossian's starting point is to point to a distinction between a personal and a sub-personal level of following rules (he does not say much about this distinction at the outset, save for offering two contrasting images: first, a person following a rule; and second, a computer following a rule: Boghossian 2008a, 479). Boghossian asserts that although there is something common at both levels, the kind of rule-following at the personal level is 'richer' than at the sub-personal level (Boghossian 2008a, 480), and he proceeds to articulate what he means by 'richer'. At the personal level, Boghossian introduces a distinction between following a rule and conforming to it, and asserts that the second is neither necessary nor sufficient for the first (Boghossian 2008a, 480). He has two examples of the necessity point, both of which are not convincing. In one, he says that (imagine the rule is 'If C, do A'): 'S may fail to recognise that he is in circumstance C, and so fail to do A; yet it may still be true that S is following R'; in the other, he says that S 'may correctly recognise that he is in conditions C, but, as a result of a performance error, fail to do A, even though he tries' (Boghossian 2008a, 480).[89]

It is not clear, by reference to the first example, in what sense S is following R: it rests on the assumption that in order to follow a rule, it is not necessary to recognise that it applies, and then apply it. But if I do not recognise that it applies and do not apply it, then in what sense am I following it? Perhaps what Boghossian meant here was that 'S failed to recognise C, and so failed to see that R was applicable, but that nevertheless he conformed to R' (but then he owes us an explanation of conformity), or perhaps by 'failing to recognise R, and so failing to do A', Boghossian meant failing to conform (but then it is unusual to think of conforming as requiring recognising that R is applicable). One of the difficulties here is that Boghossian is not speaking about conformity at all, but rather violation. Then the first example reads, on the most plausible reading, 'S failed to recognise C, and so failed to do A (that is, violated R); yet it may still be true that S is following R' – but this too is mysterious, for it is unusual to require S to fail to recognise that C (and it also

---

[89] In critiquing Boghossian's way of divorcing conformity from following a rule we are not suggesting that the two ought to be combined, as if conformity could explain rule-following. In fact, it has been argued above (in Chapter 5) in relation to appropriateness, that appropriateness is not explicable by compliance. But the way this was done in that Chapter is not to say that there is some way of following a rule that is different to complying with it, but more that neither conformity nor following a rule is either necessary or sufficient for appropriateness. In this Section, we are concerned with trying to understand how Boghossian describes rule-following, and not in making claims about the distinction between conformity and rule-following.

leaves mysterious the sense in which S, in the absence of recognising that C, is following R).

The second case is also difficult: first, it too is about violation and not conformity. Second, in what sense is trying, but failing, a violation? There is a difference between rules that require one to try and those that require one to achieve some result. Boghossian must mean that the rule 'If C, do A' is one of the latter: it does not matter if one tries to follow it – in order to conform to it, one must actually achieve the outcome it requires. But then in what sense is S following R? If R is a rule that requires outcome, rather than attempt, then S is not following R, but R* that says: try to follow R.

In terms of sufficiency, Boghossian's example is that 'Conformity to R is not sufficient for S's following R because for any behaviour that S displays, there will be a rule – indeed infinitely many rules – to which his behaviour will conform. Yet it would be absurd to say that is following all the rules to which his behaviour conforms' (Boghossian 2008a, 480). The problem here is that Boghossian is essentially talking about two different rules: ones that I can be guided by, and others that are used to describe my behaviour. It only makes sense to ask whether my conforming to a rule is sufficient for following it if it is the same rule, but then the question would be different: it would be what is the difference between a following-procedure and a conforming-procedure, e.g. must I be aware that I am following a rule in order to follow it? Thus, it is irrelevant to bring in the possibility that my behaviour may conform to other rules, for the question needs to be whether it is enough for me to be in conformity with this rule for me to also be following it (but this is to say nothing but repeat the assertion – rather than give an argument – for there being a difference between conformity to and following a rule).

Boghossian's next distinction is between following a rule and being subject to being evaluated on the basis of that rule, and here he argues that one can be subject to being evaluated on the basis of a rule without needing to have followed it (Boghossian 2008a, 480). At first blush, this seems right, but then consider what it is saying: I can evaluate your behaviour as failing to have followed a rule (when you ought to have) even if you fail to follow the rule. Is this not just saying that I evaluate your behaviour negatively when I evaluate that you did not, but you ought to have, followed the rule? Consider Boghossian's example: Nora is playing roulette and bets (all of her money), on the basis of a hunch, that the next number will be 36

(Boghossian 2008a, 480-1). We evaluate her behaviour as 'irrational' on the basis that she made her decision on 'no good evidence' (Boghossian 2008a, 481), i.e. we evaluate her behaviour negatively because we make the evaluation that she ought to have, but failed to, follow the rule that one ought to base one's decisions (when playing roulette) on good evidence.[90] But then her not following a rule is a condition of my applying the rule to her, precisely because I evaluate her on the basis of that rule. Perhaps Boghossian is aware of this difficulty, for he says that Nora need not be 'following any rule – perhaps this was just a one-time event' (Boghossian 2008a, 481). Quite aside from the question as to whether we can follow a rule on one occasion only, we can say: it is not necessary to require someone to follow a rule (any rule at all) in order to apply a different rule as a standard against which to evaluate her behaviour. But there is something that we can salvage from Boghossian's thought here, namely, that when I evaluate someone negatively on the basis that she failed to follow some rule, I am implying that I think that the rule was relevant, such that there is a sense in which I think she ought to have taken the rule into consideration, or more strongly, made her decision on the basis of the rule.[91]

So far, Boghossian has not told us anything about what following a rule consists in, so let us dig deeper. In order to follow a rule, says Boghossian, we must bear some 'appropriate relation' to the rule (Boghossian 2008a, 481). Boghossian refers to this relation as one of 'acceptance or internalisation of the rule' (Boghossian 2008a, 481). Put more clearly, in order to follow a rule I must be conforming to it *because* I have accepted or internalised the rule. Things can get complicated here very quickly, so let us look carefully at this idea. It is not clear what 'conforming to it' means here. I can say about you that you are conforming to the rule of gravity, and you can conform to it without having accepted or internalised it, but then it is not those rules that Boghossian is speaking of. He is speaking of rules such as the rule that I ought to base my decisions on good evidence. But it is a quality of these rules that they cannot be merely conformed to. Instead, these are rules that either I can use to guide myself and evaluate my behaviour, or they are rules that I can use to evaluate

---

[90] Actually, when one puts matters this way, the case becomes very strange because surely roulette is the sort of context where the rule is not applicable – it is a context in which we are expected to guess, unless we are cheating! So if Nora was to base her decision on good evidence, we might suppose she has access to some knowledge she ought not to have.

[91] But, as noted in the preceding footnote, this does not make Boghossian's example a good one – a better example would be Nora proposing some diet to a friend without having tested whether it works or checked out the reliability of the original source of the recommendation.

you. So what Boghossian is saying here is that I can guide myself and evaluate my behaviour when I accept or internalise the rule, or I can evaluate your behaviour when you accept or internalise the rule. The point is that adding 'conforming to it' does not really help.

Does the 'because' help? If 'conforming to it' does not, and the 'because' relates the 'conforming to it' to the acceptance or internalisation, then it is hard to see how 'because' would help. This 'because', it should be added, as used by Boghossian, is not a 'causal' because, but rather a 'rational action explanation.' The example he works with is the following rule (which he calls the 'Email Rule'): 'Answer any email that calls for an answer immediately upon receipt' (Boghossian 2008a, 481). Adding this to the rational action explanation of 'because', we get: 'I follow the Email rule when my acceptance of that rule serves as my reason for replying immediately, when that acceptance rationalises my behaviour' (Boghossian 2008a, 482). Notice, first, that we have lost sight of any person making an evaluation, and we are speaking only of the first-person case. Second, it is not clear what could be meant by 'acceptance rationalises my behaviour', for presumably acceptances by themselves are not the sorts of things that do rationalising (it is persons who rationalise their own or others' behaviours) – so really, we are left with the first part of the sentence, namely that 'my acceptance of that rule serves as my reason for replying immediately.'

Let us look at this more carefully: I accept the Email Rule at time T. At time T1, an email arrives. I evaluate the email as one that calls for an answer. Then I reply. Now Boghossian says that my acceptance of that rule serves as my reason for replying immediately. But what does it mean exactly for an acceptance to 'serve as my reason'? Perhaps it means that before I reply, I think to myself: 'Remember the Email Rule. Remember that you accept the Email Rule. Given that you have evaluated that you are in circumstances in which the Rule applies, apply it – follow it!' Here, acceptance adds something, but is it something that tells us what following a rule consists in? It seems to tell us something about why I might be motivated to follow the rule, and perhaps why I remember the Rule, but it does not really tell us what following the rule consists in. That, it turns out, depends on what we mean when we say that the rule 'serves as my reason' for doing something. Perhaps the best we can do is to suggest that acceptance is a pre-condition of following a rule; that before

I can follow a rule, I must accept it – but that of course does not yet tell us what following a rule consists in.[92]

Can my taking the rule as 'serving as my reason' for doing something be the same as what Boghossian means by rationalising? In the other words, do I follow the rule just in case I rationalise my behaviour on the basis of that rule? Perhaps 'rationalising' is the broader term, for I could rationalise my behaviour *ex ante* or *post facto* (before or after the behaviour). So 'rationalising' cannot serve as an explanation of what it means for a rule to serve as my reason for doing something – it is just another way of putting the same point (but in a more ambiguous fashion, for it could mean either *ex ante* or *post facto*). Taking a rule as my reason for doing something, then, is the same thing as rationalising *ex ante* my behaviour (after all, I could rationalise my behaviour *post facto*, but not have been taking the rule as my reason for doing X at the time of doing it). Of course, this does not yet help us much, for we still need know either what it means to take a rule as my reason for doing something, or what it means to rationalise my behaviour *ex ante* on the basis of that rule. Notice that all this means that it makes no sense to say, as Boghossian says, that 'If S is following rule R by doing A, then S's acceptance of R explains S's doing A', or that it 'rationalises S's doing A' (Boghossian 2008a, 483). S's acceptance, as noted above, may help us understand why (in a causal sense of why) S remembers R or is motivated to remember R, but not in what sense S is following R. Further, it makes no sense to have 'rationalising' on the opposite side of the definition, for we have just said that one way of articulating what it means to say that S is following R is that S rationalises her behaviour *ex ante* on the basis of R – and it is not informative to say that what it means to rationalise my behaviour *ex ante* on the basis of R is to rationalise my behaviour on the basis of R.

Perhaps what Boghossian wants to say is that my having accepted the rule means that when I follow it, i.e. when I take as my reason for doing something, I am also going to believe that I ought to do what the rule requires me to do. But this does

---

[92] Another way of saying what we want to resist here is the regularisation of acceptance, for once one regularises acceptance one is on the way to saying that we are perpetually – by default – guided by the rules. If we regularise acceptance, we make it out as if the rule is always relevant – our point being that that is not so, for we are not passive in this way vis-à-vis relevance (what we need to avoid, in order to understand the mind better, is precisely avoid this passive positioning). The regress problem is not something to be solved: instead, it shows us why we cannot regularise the relevance of normative resources (they are contingently useful, and not necessarily applicable – whether *a priori* or by defeasible default).

not seem to add anything that the concept of a reason does not already say, i.e. if I take something as my reason for doing it, I am taking it as telling me what I ought to do. The point is that my prior acceptance of a rule does not adding anything to explaining what it means for me to take that rule as my reason for doing something.

There is an important consequence in taking rule-acceptance as an explanation / rationalisation of me taking a rule as a reason for doing something. It positions the regress problem at the level of rule-acceptance. Thus, a large part of Boghossian's discussion revolves around the issue of how I can accept something without running into the regress problem. It turns out, for instance, that saying that I accept something 'intentionally' just pushes the problem back one step, for I now need to say what acting intentionally consists in, and I come to posit (for what else?) another intention to act intentionally, and so on. That is why, says Boghossian, we end up with the idea that there must be, at bottom, some kind of non-intentional rule-acceptance (Boghossian 2008a, 487). As he puts it: 'to rescue the possibility of rule-following…we must find a way of accepting a rule that does not consist in our having some intentional state in which the rule's requirements are explicitly represented' (Boghossian 2008a, 494). But the question that we can ask here is why does Boghossian think that characterising rule-acceptance will offer not only an explanation / rationalisation of rule-following, but also save it from the regress problem? Why does the regress problem occur not at the level of actual rule-following, but at the level of rule-acceptance? Why does Boghossian think that when Wittgenstein speaks of 'blind rule-following' or 'obeying the rule blindly', he is speaking of some non-intentional form of rule-acceptance? Why does Boghossian not try to find the answer at the level of rule-following – indeed, why does he not give us an account of rule-following, but only one of rule-acceptance?

There are two observations to make about these questions. First, it is not quite accurate to say that Boghossian does not say anything about rule-following. To see that he does say something, consider the following example (referring back to the Email Rule). The discussion of the example occurs in a context in which Boghossian takes himself to be speaking about the problems that arise for an intentional reading of rule-acceptance, but it pays to look at it here for in fact it tells us something about the process of rule-following. Because he is talking about intentions, Boghossian first translates the Email Rule into what he calls 'an intention': 'For all x, if x is an email

and you have just received x, answer it immediately' (Boghossian 2008a, 492). He then says this:

> To act on this intention, it would seem, I am going to have to think even if very fleetingly and not very consciously, that its antecedent is satisfied. The rule itself, after all, has a conditional content. It does not call on me to just do something, but to always perform some action, if I am in a particular kind of circumstance. And it is very hard to se how such a conditional intention could guide my action without my coming to have the belief that its antecedent is satisfied. So, let us imagine, then, that I think to myself:
>
> > Premise: This is an email that I have just received.
>
> in order to draw the
>
> > Conclusion: Answer it immediately!
>
> At least in this case, then, rule-following, on the Intention model, requires inference: it requires the rule-follower to infer what the rule calls for in the circumstances in which he finds himself. (Boghossian 2008a, 492)

Now the interesting thing about this passage is that, absent the language of intentions, and absent the fact that Boghossian thinks he is talking about rule-acceptance, the passage helps us understand more of the process of following a rule. It still leaves some things unexplained, but it helps. It does so because it points to an inferential process involved in seeing some set of circumstances as the antecedents of the rule. So we have a clue here: rule-following consists in taking the rule as a reason for my action, which means engaging in an inferential process whereby I see some set of circumstances as the antecedents of the rule, and this triggers the consequent of the rule that I then use as my reason for acting. Of course, we still need to explain what is meant by me taking the consequent of the rule as the reason for my acting. But it would be going too far to say that nothing is explained about rule-following, for this process of inferring that this set of circumstances fits the antecedent of a rule is properly part of a description not of acceptance of the rule, but of following it. There are still questions that we can raise about what inference is here, e.g. what does Boghossian mean when he says that 'I am going to have to think even if very fleetingly and not very consciously, that its antecedent is satisfied' – what kind of 'thinking' does Boghossian invoke here? But this is a start, and the interesting thing about is that it not only helps us understand better the process of following a rule, it

also (and this is crucial) helps us to see that we cannot possibly be following rules all the time (as if by default), i.e. if the process of following a rule does require me to see some set of circumstances as the antecedents of a rule (i.e. to infer that these circumstances are also the antecedents of the rule), and if this process requires some act of 'thinking' (to use Boghossian's term) on my part, then this helps us see that it is implausible to suggest that the mind can only be a rule-following thing, or that it essentially or necessarily is. This is because any reflection on our daily practices reveals that we do not always – indeed, not even often – engage in such inferential processes.

Boghossian would not find this way of proceeding satisfactory, and this brings us to the second remark to make about the above questions. It seems that Boghossian wants to place the emphasis on rule-acceptance, rather than rule-following, and to insist on the possibility of non-intentional rule-acceptance, because he does not want to restrict rule-following to just one way in which the mind can learn and function, but as the necessary and essential way. The point is that it is difficult (perhaps impossible) to make sense of non-inferential rule-following that can occur without 'thinking' (which would, as noted above, restrict its reach) – and that is why Boghossian places the pressure on rule-acceptance to be non-inferential / non-intentional. After all, to follow a rule (on Boghossian's own account) does seem to involve going through an inferential process (which involves some act of thinking) of seeing the circumstances as antecedents and then acting on the consequent associated with the antecedent (where 'acting on' still needs explanation). The most straightforward way out of the dilemma is to simply restrict rule-following to that inferential thinking process (though with the proviso that many things are still left explained), but Boghossian does not want to do this, for this would radically restrict the applicability of rule-following as an understanding of how the mind learns and works.

Now all this raises an obvious question: but are you saying there really is no blind rule-following? And how could you maintain such a position: what about me following the rule for addition – surely, I do not 'think' (that word again) about the rule every time I employ it? This is another form of the question we are tackling in this Chapter: what we should say here is that the question forces us to think about the mind as a rule-following thing, trying to squeeze our description of how the mind might learn and exercise the ability to add by describing that learning and any act of

adding as one of rule-following. But the very point is this: that perhaps learning and following rules is not the way the mind learns to, and does in fact, add. Of course, this in itself does not mean that no rule learning and rule following is possible: all it means is that this process might be much infrequent than we imagine is the case. And, what has been said above about Boghossian is that he helps us, somewhat unwittingly, to see this, for he shows us that the process of rule-following is an inferential, thinking process (or put differently, precisely by moving the analysis to the second-order level of rule-acceptance, and trying to find a solution to the regress problem there, Boghossian shows that it is difficult, if not impossible, to understand rule-following in any other way).

Let us now continue, and move on to another prominent attempt to 'rescue rule-following', while keeping one eye on the reasons for which theorists might be attracted to rescuing it, and how we can see – instead – that it is in fact a limited account of how the mind learns and works, and that there is much that is missed when we focus only on rule-learning and rule-following.

C. Pettit's Inclinations

The next prominent attempt to rescue rule-following that it will be worthwhile to discuss is Phillip Pettit's 'The Reality of Rule-Following' (1990). Pettit argues three things:

> Firstly, that the fact that any finite set of examples instantiates an indefinite number of rules does not mean that it cannot exemplify a determinate rule for a given agent; secondly, that the set of examples can exemplify such a rule if the examples generate an inclination in the agent to go a certain way: the rule exemplified will be one which is suitably associated with the inclination; and thirdly, that a suitable association between inclination and rule is this: that the rule is that rule to which the inclination corresponds in the actual world, provided the inclination operates under favourable conditions. (Pettit 1990, 14)

We shall not analyse here what properties Pettit attributes to rules; rather, our focus is on what picture of the mind Pettit draws for us on the basis of his pursuit of the 'reality of rule-following' (our interest, then, is primarily in Pettit's notion of an 'inclination').

That Pettit thinks a lot is at stake in this debate is clear from his claim that without the capacity to learn and follow rules, there would no thinkers and speakers

(Pettit 1990, 1). He qualifies this somewhat by saying that this does not mean there could not be 'intentional and even social' subjects (Pettit 1990, 1), but this offers little comfort. Pettit also characterises the sceptical challenge (to rule-following) in very strong terms, i.e. it is an all-or-nothing matter: either there really is rule-following (at which point there is also thinking and speaking) or there is no rule-following (at which point we concede that persons are not quite what we have for a long time thought they have been, i.e. thinking and speaking animals). It is not clear why Pettit thinks that the stakes must be so stark, i.e. why it could not be the case that sometimes when we speak and think we guide ourselves by rules, and sometimes when we speak and think we do not do so. For Pettit, however, the matter is stark. Further, he asserts that it really is the case that we understand ourselves as – and that we really are – rule-following animals. This claim is not so much argued for, as it is asserted. Thus, in the case of speech, Pettit says that 'it is generally conceded that we are required to be able to follow rules so far as we speak' (Pettit 1990, 4). No such thing is generally conceded, though perhaps in 1990 – at a time when Chomsky dominated linguistics, and in a time when connectionist accounts of language learning and use had not yet reached their peak – this was a more defensible assertion.

Pettit thinks there is more to argue about in the case of thinking. Thinking, he says, requires more than 'just the having of intentional attitudes' (which Pettit describes as attitudes 'of belief and desire': Pettit 1990, 4); the requisite intention needs to be 'produced in each case…in the right way…by the desire for a certain state of affairs and the belief that doing this or that offers the best promise of desire-satisfaction' (Pettit 1990, 4-5). On this view, our intentional attitudes have to be shaped (produced) with a view 'to having beliefs that are adequate for certain projects, or beliefs that are true' (Pettit 1990, 5). Yet again: 'The thinker must be able to wonder whether something is so, to institute tests to see whether it is so or not, to accept in the light of those tests that it probably is, and so on' (Pettit 1990, 5).

There are two observations to make about this. The first is that Pettit's model of a thinker is very demanding – perhaps on an analogy with an experienced, responsible and almost ideal researcher. For Pettit's thinker is someone perpetually involved in a project, for which certain means are laid down as paths to follow, and who is ceaselessly operating in a context of justification: making claims (that something is so-and-so) and not only supporting this with evidence, but also testing that evidence to verify his claim. The second observation is to query whether, even if

our model of thinking is so demanding, where we can only be responsible in the way that Pettit suggests we can be, i.e. whether, for instance, aiming at the 'adequacy' of beliefs, or producing intentions in the 'right' way, must mean following rules. At stake in the second observation is precisely the understanding of the normative, but we also need to take note of the first observation, for seeing that we are not confined to thinking in the way Pettit thinks we are helps to see how the normative is misunderstood when it is thought that adequacy and rightness are determined by rules.

To help put some more meat on the bones of the first observation, consider the one example Pettit offers of thinking: a 'thinker who wonders what is the sum of two numbers' (Pettit 1990, 5). Pettit says that this 'wondering' is exactly the same process as the process that a 'speaker who sets out to apply the word "plus" properly' goes through (Pettit 1990, 5). But is this a plausible understanding of 'wonder' or 'wondering'? Wonder is an interesting term to use, for we might think that wondering is precisely unlike calculation (as Pettit suggests it is); there is, is there not, a difference between 'wondering' about the sum of two numbers and 'calculating' that sum? What, then, might it mean to wonder at the sum of two numbers? Well, for one, it might mean considering what different ways there might be at getting at the sum, e.g. if the two numbers are 42 and 67, one might wonder if it is going to be easier to get at the sum if one breaks this down into, say, 40 and 60, and 2 and 7 (the point is that this is not calculating, but wondering about how to calculate more efficiently)? Or it might mean wondering just what else 'plus' might be – what else it might be to put things together – e.g. imagine we wonder whether it would be better, when we give an answer to what 42 plus 67 is, to say not 109, but 90-and-19 (perhaps there is some advantage to be gained in making the number 90 so prominent)? Perhaps even these examples of wondering are somewhat tame for they do not actively challenge the aim (of plus-ing). Perhaps, then, we need to make room for the possibility of wondering just what 'plus' is about, what purpose it serves, how it could be otherwise (could it, for instance, be divided into two different rules?).

The general point is this: what Pettit seems to have in mind when he thinks of 'thinking' is some form of calculation, which has a set end and a set means for getting there, and which involves performing or enacting this procedure – or, perhaps, more generously, some sort of process in a context of justification, where we aim to make a justifiable claim about how things are or should be (or what we should do). Pettit's thinker is perpetually engaged in the pursuit of a project (e.g. adding numbers) for

which (and this is a crucial assumption) there is a rule that dictates the proper means for achieving the aim of the project (the rule for addition). But does thinking always, and by necessity, have an aim in mind, and even when there is an aim, is that aim one that is always, and by necessity, governed by some means that the thinker has to recognise and apply in order to achieve that aim? If we think thinking is like this, we miss out, surely, on a great deal of rich mental states and processes (we have referred a number of times above to appreciation, but we can think of many others, e.g. wonder, admiration, puzzlement, uncertainty, bewilderment, and others) about which it would be controversial to assert that they were thoughtless. Of course, one can define 'thinking' narrowly, but then when one combines this with the idea that what is at stake is our very understanding of ourselves, one risks painting a very narrow picture of how the mind learns and works.

In light of the above, one might already legitimately express concern about the argument for the reality of rule-following (as set up by Pettit). For if one begins with the postulate that thinking and speaking are necessarily rule-governed, one puts oneself – as Pettit does – in the position where one is forced to try to find a way for the mind to always and necessarily be learning rules and following them. This not only leaves no room for an account of learning and for abilities / dispositions / inclinations (or whatever one calls them) that are not dependent on rules; it also might make one's account of rule-following so broad that one loses sight of those times in which the mind might legitimately be said to have learnt a rule such that it can follow it. But let us, all the same, look at Pettit's notion of inclination as his answer (or at least the part of the answer that interests us most here) to the rule-following sceptic.[93]

On Pettit's picture, an inclination emerges from how a rule comes to be 'represented to a human subject' (Pettit 1990, 9). It does so, Pettit argues, through a series of 'examples of its application' (Pettit 1990, 9). Imagine, then, that we are faced with the following examples: '(1, 1, 2)-(1, 2, 3)-(2, 2, 4)' (Pettit 1990, 9). Clearly, Pettit cannot simply assert that being faced with these examples just means that 'the' addition rule is represented to us, for the sceptic would be quick to respond (as Pettit well knows) that the examples give ample (indeed infinite) possibilities for articulating some rule (they can be made out to accord with an infinite many rules).

---

[93] Again, we should stress that our position is not a sceptical one: Pettit makes the debate too stark, as one between the reality of always following rules or never following them; instead, it is being argued here that a position in the middle is possible.

So how is it, then, that the examples represent that particular rule, and no other, to the subject? Pettit's first step here is to say that what we are concerned with here is not the one-many relationship of 'instantiation' (where indeed one set of examples can instantiate many rules), but the two-place relationship of 'exemplification', namely one that 'involves not just a set of examples and a rule but also a person for whom the examples are supposed to exemplify the rule' (Pettit 1990, 9). That is fine, but it does not get us far, for the issue precisely is how / why the person is presented with just that rule and no other.

The second step is to suggest that 'an agent develops an independent disposition or inclination to extrapolate in a certain way to other cases: an inclination of which he may or may not be aware' (Pettit 1990, 10). What this adds to our puzzle is that I may learn a rule without being aware what rule I am learning – but it says nothing about why it must be that particular rule and no other (in fact, there is more to this step, for it suggests that whenever I am exposed to a series of examples, I must develop a disposition or inclination that takes the form of a rule – this being a supposition that we will question – but for the moment let us ignore it).

It is the third step that Pettit presents as the crucial one. What the third step introduces, in effect, is a sense of my being able to tell that I have gone wrong (though not wrong with respect to the rule, but wrong about the rule exemplified in the cases). Pettit asks: 'What…way is there for a rule to relate to my inclination?' and he answers:

> It can only relate as that rule which fits my inclination but only so far as certain favourable conditions are fulfilled: in particular favourable conditions such that I can discover that in some cases they are not fulfilled, and that I got the rule wrong. The rule associated with the inclination will be that rule, the one that satisfies this inclination, provided the inclination fires under the conditions identified. (Pettit 1990, 12)

This does help, but only to the extent that one assumes that the future examples will be examples that are governed by some particular rule, for how else could I 'discover' that I got the rule 'wrong'? The image Pettit has in mind here is my learning some rule from the examples, and then, when faced with extra cases, thinking to myself (or at least having the capacity to think to myself): but that was the wrong rule to learn. That is fair enough, but what guarantees that the rule that I then take myself to have learnt (the new rule) is the one and only rule that is available to be learnt (on the basis

of the now enlarged pool of cases)? Pettit is right that I can have the experience, having learnt a rule from some examples (say 3 examples), of being struck upon seeing another set of examples (the second 3, so examples number 4, 5 and 6) that I should not have learnt that rule but some other rule – but what he does not tell us is why the new set of examples – now six in total – makes it the case that I must learn just that particular rule and no other one (as if there really was only one rule to learn from the enlarged pool of examples).

Notice that Pettit makes a very valuable observation: he does not say that when faced with the next set of examples, I will automatically and unrevisably think that they must either exemplify the rule I learnt or some other rule (because they do not fit the one I learnt); he allows for the possibility of me revising the rule I learnt to cover the new set of cases. But the problem faced by Pettit is still the same one with which he began, namely: why is it that particular rule, and no other, that I learn? In the absence of any argument, we must reply: there is no guarantee that I will learn any particular rule, for it is not necessary to learn any one particular rule (this does not mean that I cannot take some set of examples to exemplify a rule – it just opens the possibility of saying that I need not commit to any one particular rule when learning). For Pettit, human beings pick up rules (in the form of inclinations) from examples, and not only must they, but they only ever can, pick up one rule (*the* rule) from those examples (unless they make a mistake, which they can 'discover' later).

Here, we can make three observations: first, there is the issue of what ought to count as an example, and which cases will be seen as part of the same pool of examples and which will not. Recall that in his analysis, Pettit makes room for the possibility that when confronted with a new set of examples, we will not use the rule that we learnt to force the new examples to comply with it (or dismiss the examples as irrelevant, and thus not engage with them at all), but rather that we will consider revising the rule, thinking that perhaps we got the rule wrong. The question is: what makes us see the three new examples as relevant – as part of the same pool as the first set of examples? And what makes us see the initial cases as examples of one rule, and not three different examples? Asking these questions also raises the familiar chicken-and-egg problem: is it the rule that comes first, for it groups things into examples, or is it the examples that come first and the rule just one way of interpreting them – but if the second, then how come I see things as examples? These questions receive no reply in Pettit's analysis. What they do help us to see, though, is the limits of a rule-

based analysis – for, if nothing else, a rule-based analysis (one that sees us as learning a rule from a set of cases) does not tell us how we see things as cases ripe for exemplification (the point is to see that we are active and creative in this respect, i.e. that we are not confronted with what already are examples – we must treat them as so, meaning that we may also not, i.e. we may treat them as not usefully groupable).

The second observation to make is why think that we must learn only one rule from a set of examples. It is possible, certainly, that when faced with a series of (what we take to be) examples, we will come up immediately with one way of relating (one rule that relates) them to each other, but it is also possible that we may come up with a variety of rules and conclude that we cannot know which ones to keep alive and which to exclude (as possibilities) until we enlarge the pool of examples (here learning is already something more tentative: to learn we do not need to commit; indeed, we may precisely need to not over-commit).

The third observation to make is why call what we are learning, even on Pettit's own account, 'rules'? Pettit's analysis can point either to a way of interpreting what is presented to us as a set of examples, or as a way of seeing things as examples; in both cases, we might be able to articulate this in the form of creating and thereafter being able to recognise a pattern. We need not be scared of the word 'pattern', i.e. we need not think that when we create and thereafter are able to recognise a pattern that we are somehow not learning how we 'ought' to go on – for whether we call it a rule or a pattern, it is likely that we will have no sense of learning how we 'ought' to go on until we criticise ourselves or are criticised by others.[94]

We are not necessarily endorsing talk of 'patterns' here, but the point is that Pettit's notion of an inclination developed in the course of its encounter with examples (which is an interesting notion) need not have anything to do with rules. What we might be seeing here is how, if we start from the ground up – of things appearing as examples, and of us creating ways of seeing patterns in examples, and thus not presupposing rules – we can do so without an explanation that refers to rules. Of course, this need not mean we should get rid of rules altogether, for sometimes we learn by being told that there is a such-and-such a rule, and that it is exemplified in

---

[94] When we criticise ourselves, we might thereby turn a pattern into a rule – we, as it were, commit more strongly to that grouping. In the case of being criticised by others, it is important to make room for indeterminacy, e.g. it does not follow necessarily that if others tell us that we saw the 'wrong' pattern that we will immediately comply and come to see things the way they do – we may, or we may not.

the following cases. But what Pettit shows is that if we begin with the encounter of a mind with the environment, we allow for the possibility of a picture where the mind is much more active – it creates a pattern that allows it to orient itself tentatively, and in a way that is not closed off to the possibility of revising the pattern (where, furthermore, the pattern may be transformed, by the person, into a rule as a result of criticism).[95]

We are trying, here, to read Pettit generously, for Pettit's own cases are ones where he presents us already dealing with examples (his descriptions assume for us what counts as an example) that, furthermore, exemplify one rule, i.e. the way he sets up the case is that I am given (passively) a set of examples that (he assumes or posits) exemplify a certain rule, and the issue is whether I will learn this rule (correctly recognise it as exemplified) – e.g. he says, 'I can intentionally conform my behaviour to a rule exemplified for me by certain examples' (Pettit 1990, 15). The point being made here is that neither are the examples, nor is the rule, made *for* me, but *by* me.

In effect, Pettit begins with an ontologically stable and secure order of examples held together by certain rules. The epistemology then follows: there is this set of examples, which exemplifies this one rule, and the question is whether, and how, I will learn it (and thereby act correctly). This positions us passively before phenomena imagined by the theorist to be a certain way; in other words, here the pursuit of the normative (as a rule-governed domain) makes us fix things (fix what is there to be learnt / seen) with the result that we are pictured as passive receptacles (if we are appropriately obedient!) of what is said to be there. The way we have tried to read Pettit above, though, opens up the possibility of setting up the case in a different way, i.e. not beginning with a metaphysically stable case, but beginning with an epistemological picture according to which the mind is active and creates how it (tentatively) groups things or treats them as patterned.

Reading Pettit this way is a step forward, but we can, and ought to, go yet further still. Recall the above suggestion that what we may be learning may not be one pattern, but coming up with certain hypothetical possibilities, and suspending our commitment to any one until we are faced with more examples. Appealing to a pool of possible patterns is already an improvement, but perhaps we can build on this and

---

[95] This need not be confined to criticism by others, but may include criticism of myself by myself, as when, e.g. I castigate myself, after the fact, for not relying on that pattern (thereby turning it into a rule for myself).

say that it is also possible that we experience things in such a way that acquires resources that may be grouped, in the future, in a variety of ways; those resources, in other words, could come to be seen (in light of future experience) as sets of examples, even though at the time we experience them we need not experience them as examples (of course it is also possible, and it is important to see, that we can actively and creatively treat things as groups of examples). It may be that this sense of experience is illustrated well in works such as Marcel Proust's *In Search of Lost Time* (1913-1927), who often first describes how something was experienced, but only in later passages – in light of some new, often quite accidental experience – does the person come to see those earlier experiences as part of a series of examples of something (resulting in, for instance, some observation about his own character). This would suggest our experience is much more supple and flexible than we think, and indeed that what we take ourselves to have experienced is itself subject to what, in the future, we take ourselves to learn about our past. What one could say here is that although I can learn something specific, e.g. some rule or a pattern, or even a group of rules or patterns, my learning experience entails more than that, where the 'more' remains indeterminate, and thus open to being construed differently in the future.

There is more, of course, that would need to be said about this possibility, and about learning in general. Unfortunately, such a theory of learning lies outside the scope of this thesis. The point, in this context, is simply to raise it as a possibility, and thus to show that this possibility is missed when we set up cases in such a way we assume or posit we are always and already presented with examples in which there is one and only one rule to be learnt, and then go on to describe how we learn that one rule (or failed to learn it, but then discover that we were wrong and learn the correct rule after all). In other words, when we focus exclusively on learning and following rules (and not just any rules, but what we assume, as the theorist, to be *the* rules) we miss, a great deal about how actively and flexibly the mind learns and orients itself.

D. Stueber's Dispositions

We have seen that Pettit refers to inclinations (though sometimes also to inclinations as dispositions). Others speak precisely of dispositions. One theorist who does so is Karsten Stueber (who we encountered briefly above in Chapter 2), though his account of dispositions (in his paper, 'How Think about Rules and Rule Following', 2005) is qualified in a way that will help us clarify some of the issues in this context. Stueber,

as we shall see in a moment, refers to 'complex dispositions', which involve both first-order dispositions to comply with rules, but also second-order dispositions to engage in a critical reflexive attitude, i.e. an attitude that uses those rules as standards for the evaluation of one's behaviour and that of others. He argues that where others have gone wrong is in thinking either that the acquisition and exercise of such complex dispositions has to be computational in character, or in thinking that because it has to be computational in character, so we must do without rules and speak only of non-rule governed abilities (e.g. embodied know-how and the like). In other words, Stueber thinks rules and rule-following ought to be saved, and can be, but only if we accept the idea of complex dispositions that are non-computational.

Before we look at his account, we need to notice a number of assumptions that Stueber is working with. The very first paragraph of the paper conveniently summarises these assumptions:

> It seems to be a platitude to point out that the social realm is the realm of rules whereas nature is the realm of natural laws and brute regularities. Only the behaviour of human agents, especially in a social setting, is bound and guided by rules, and because of this feature their behaviour can be normatively evaluated as right or wrong, appropriate or inappropriate, and rational or irrational. It is part of the education of every child at home and in school to be taught the rules of the 'game' in order to become a productive member of society. Social scientists, historians, and anthropologists are thus prima facie well advised to study the normative standards of a particular society because it is in light of these standards that they are able to explain why individual members of that society act the way they do; why they drive on the right side of the road instead of the left or why eat with knives and forks instead of their hands. (Stueber 2005, 307)

We have already seen, in both Part I and II, how the divide set up between the normative and the natural (or the normative and the regular) often leads to thinking of the normative in a very narrow way, e.g. as a matter of rules and rule-following (where the problem, ironically, is that the governance view goes on from there to regularise rules and rule-following). We have also seen that it is unnecessary, and indeed misleading, to think that something cannot be appreciated as right or wrong, appropriate or inappropriate, even rational or irrational (though we noted it may be better to opt for the more flexible concept of reasonableness and not rationality), without being determined by rules (or other alleged determinants of appropriateness and inappropriateness). We have also seen that prioritising rules in this way can be

self-serving, for it can produce examples (or targets / objects of explanation) that appear to be inexplicable in any other way than on the basis of rules (i.e. as we saw in Chomsky, social regularities are defined in terms of certain rules, such that it becomes difficult if not impossible to see how social behaviour could be explained in any other way than on the basis of how persons learn and do follow rules). We shall not rehearse these claims here – but it is useful to point out that Stueber's analysis presupposes them (insofar as Stueber speaks at the social level, some issues have not yet been dealt with and are so in Part IV of the thesis).

Stueber's first step is to criticise the need to require persons to be guided by a rule only when representing the rule to themselves and acting on that representation. He criticises this move irrespective of whether it suggests that representation needs to be conscious on every occasion, or whether it opts for some form of 'internalisation', e.g. by 'social conditioning and training or because of our innate psychological endowment given to us by Mother Nature' (Stueber 2005, 310). He calls on the model of a computer – 'a symbol-crunching machine' (Stueber 2005, 311) – to tell us that internal representations are not necessary for rules to 'still play a causal role in the production of behaviour' (Stueber 2005, 311), but he thinks that the computer model still falls short because it fails to meet three problems: first, the normativity problem (by which Stueber means the way that the rule can serve as a standard for the evaluation of behaviour); the frame problem, i.e. accounting for how rules are relevant / applicable; and third, what he calls 'the methodological quandary', by which he means that 'we could come up with different systems of rules that describe a particular practice adequately' (Stueber 2005, 314), such that we are simply assuming that certain rules are being internalised by the agent, but have no grounds for this assumption.

According to Stueber, whereas the computational model of rule-following runs into these problems, Stueber's proposal does not. We must look at why he thinks it does not. Stueber places great faith and much explanatory pressure on the way in which 'agents themselves accept normative evaluations of their behaviour as being appropriate, why they can rationally discuss the appropriateness of their behaviour, and why recognising a violation of a normative standard can be a reason for changing their behaviour' (Stueber 2005, 316). In other words, our talk / discourse about how we ought to behave is given a vital role in explaining how we do behave. The way that Stueber thinks the one is connected to the other is via mechanisms of

'monitoring' (both self and other). This monitoring is not a process that goes on all the time whenever we act, but it kicks in when we 'recognise normative violations' of rules. Thus, he says that although 'we normally are not aware of the fact that we are following certain rules', what 'makes our social skill a practice of rule following is the fact that I recognise normative violations of such a rule' (Stueber 2005, 319). The way we react (at the second-order level) to what we recognise are violations is, it seems for Stueber, proof that we were in fact following such rules (at the first-order level): 'agents can be understood as following rules if and only if agents have acquired certain complex dispositions that not only include first-order dispositions of showing regularity of behaviour, but also involve reflective capacities or second-order dispositions to monitor one's response for its correctness or appropriateness. Such higher order dispositions', Stueber continues, 'are normally only actualised in a situation in which the appropriateness of one's behaviour becomes an issue. They are not ordinarily involved in causing behaviour in unproblematic cases' (Stueber 2005, 320).[96]

To get off the ground, Stueber's account would have to show what he means by saying that we 'recognise normative violations.' Notice that the way Stueber sets up his cases, what is appropriate or not is already defined by the rules; we come to learn how to act appropriately by learning the rules (though not necessarily with any awareness), and the fact that we have learnt and are following rules is only 'revealed' in the moment we 'recognise violations' of those rules. To see that it is untenable to claim that the recognition of violations reveals underlying rules, consider the following seven counter-examples:

- First, it is possible to criticise oneself for failing to be properly guided by some rule, e.g. as when someone learning to follow some rule keeps criticising himself for failing to follow it (e.g. failing to remember it and failing to apply it). Imagine that I learn to play chess by reading a rule-book, and then I come to play a game (with myself) – here, I might keep checking whether the moves I am making are in accordance with the rules. As I do so, I might say to myself

[96] Notice the many overlaps with the regular here: first, there are underlying regularly performed behaviours; second, there is a second-order disposition, which is something capable of being regularly exercised whenever there is a recognition of rule violation; and third, this second-order disposition, as a result of being said to be capable of being exercised regularly, is treated as evidence of a first-order disposition to regularly follow the rule.

that I am not acting in accordance with the rules, and criticise myself for that – it does not follow from my criticism that I was following the rules – just that I was trying to. Here, my reflective critical attitude reveals just that I am committed to learning the rules, not that I was in fact following them.

- Second, it is possible for me to create a standard of evaluation for my behaviour: e.g. imagine that when I am playing chess, I notice after a while that I take a lot of time by the way in which I take pieces (e.g. I pick up my piece, knock over the other one, put my piece on where the other one stood, then pick up the taken piece), and that I can save time by first picking up the other piece, taking it off the table, and then putting my piece there. I try this out and it does save time. Seeing it works, imagine that I tell myself that I ought to do this in the future, thereby creating a standard by which I will evaluate my behaviour when playing chess. Two things can be said about this case: first, my criticising myself for being slow in the manner in which I took pieces is not really a criticism of violating some rule;[97] second, the fact that I can create this evaluation for myself shows that we can engage in social activities (like playing chess) without all being governed by the same rules (by some 'game', as Stueber puts it, which determines the relevant ways for criticising myself and others).

- Third, it is not a given that if I do think to myself that I have violated some rule, I will also immediately criticise myself on the basis of having violated that rule, i.e. the way in which I evaluate myself (and others) is not determined even if I am committed to following a certain set of rules in some domain (one only thinks this if one thinks that appropriateness and inappropriateness are determined). For example, imagine I am committed to following and monitoring my behaviour in accordance with the above rule for taking pieces.

---

[97] One might try to invoke a value, i.e. one could justify one's own criticism to oneself that perhaps one is a person cares about being efficient, and although this is possible (and an improvement on it being necessary to invoke a rule), it would be problematic to require it by necessity (e.g. that one either, by necessity, invokes a rule or a value). The point is that one's criticism of oneself need not be revealing of anything prior – it need not be enabled by, but rather potentially bring about, commitment to some value (or a rule). Given what was said above, in Chapter 6, on inappropriateness, it is possible for one to appreciate something one does as inappropriate in a new way – not a way that must, by necessity, be an instantiation of something previously accepted.

When I am playing a game, it is perfectly possible for me to evaluate my behaviour as appropriate (despite using my old method) – i.e. I can create, as I go on, certain limits as to how much I want to be committed to the rule, or what I want to take as the scope of the rule (certainly, sometimes I may also criticise myself for being too lax about some rule, perhaps because I am really committed to learning it – but the point is that there is no determinacy here about how I will evaluate myself).

- Fourth, the way I criticise or justify myself (or the actions of others) need not by necessity map onto what I am guided by or oriented by when I act. It is perfectly possible for me not only to deceive myself, but also to want others to think of me differently than I think of myself (in this second sense, I am honest with myself, but not honest with others). Thus, imagine that I know that, at bottom, I am a racist: I think less of persons with such-and-such a skin colour, or I don't trust them. Imagine, for argument's sake, that I am guided by this in everyday life. It is perfectly possible that I will criticise others (in the company of others) to try to show to them that I am committed to the same rules they are. Here, we would not say that my 'reflective normative attitudes' are somehow proof of my following such rules. Notice that this case can probably be changed so that it is not entirely hypocritical or in such bad faith, e.g. imagine I am new to some community, and that I have just heard people talk about racism in a negative way, but that I have not yet completely understood what racism is really about (generally, or for this community specifically); I may want to be equally critical of racism, so as not to appear too much of an outsider, but yet fail to be guided in any way by the rules defining racism (of course, we are assuming that racism is rule-governed just for the purpose of debate). That is a case of me deceiving (or seeking to influence) others, and possibly something like semi-deception of myself (for I could tell myself I am committed to anti-racism, but yet not quite know what that means), but I can also deceive (or seek to influence) myself more thoroughly: imagine I justify smoking to myself by telling myself I need it to relax myself, and to stop me from being aggressive, and imagine then that I act aggressively on some occasion – I could here criticise myself for failing to smoke, whereas I am also aware (if I was to perfectly honest with myself) that

I was aggressive for some other reason (e.g. imagine I am aware, though I do not want to acknowledge, that my pride was hurt).

- Fifth, it is perfectly possible to imagine that I simply do not know why I did what I take myself to have done, such that it makes little sense – even if I am committed to using some rule as a guide and a standard for evaluation – to apply that rule to myself.

- Or, sixth, imagine that someone criticises me for being aggressive, whereas I honestly do not think I was, but I am nevertheless unhappy with myself for somehow allowing (or was it really 'me'?) such a situation (i.e. a situation in which someone could accuse me of being aggressive) to occur.

- Or, seventh, imagine that I criticise myself for not caring about something I think, on reflection, I ought to have cared about (e.g. I ought to have paid more attention to that sunset last night – I do not care enough about the beauty of the natural world) – here, it does not look like I tell myself off for violating some rule: I just tell myself to care more about things I feel I ought to care more about (it feels strange to say I am governed by a rule that I ought to care more about the beauty of the natural world, if only because I may have only felt this upon remembering how my friend was moved when he saw the sunset, and I feel sad that I did not accompany him in that feeling).

Perhaps there are other cases, but these should suffice.

It is useful to add here that there have been many studies by sociologists and psychologists that have pointed to some of the above ways in which my normative reports of myself (or others) ought not to be taken at face value. For example, the sociologist David Cheal (1980) has pointed out that 'excuses can be interesting sociological data' (Cheal 1980, 44) – 'interesting' precisely because they can mask just as much as they reveal: 'Actors may not articulate the rules they follow, nor accept accurate formulations of those rules by observers, for manipulative reasons… Even where actors are perfectly honest with the researcher, however, they may not give the real reasons for their behaviour. They may prefer a little self-deception,

because they have problems facing certain of their feelings, or simply as part of the process of constructing an orderly history of their experiences' (Cheal 1980, 46). Psychologists have long pointed to our tendencies to justify our conduct to ourselves (and often also to others) by whatever means we can. Here, then, we can conclude that positing or assuming that the practice of reflective normative attitudes maps onto the rules allegedly governing how we (allegedly) really do behave is wishful thinking, and fails to consider the unique contexts of such attitudes while also failing to see the various relationships that can exist between behaviour and such reflective normative attitudes. Perhaps it is less neat to loosen the potential ties between ways of evaluating ourselves and others and how we may be guided, but as soon as we loosen things, we also notice more (as the above examples hope to indicate) about how the mind learns and works, and about how persons relate to each other. Referring, then, as Stueber does, to complex dispositions, where these depend, greatly, on the plausibility of reflexive critical attitudes revealing prior commitment to, and / or the fact of, being guided by, rules, does not work, and it does not work because it is implausible (and somewhat naïve) to think that such attitudes must map on to what we may (on occasion) be guided by.

We have now spoken of inclinations and dispositions.[98] Let us turn next to abilities.

E. Tanney's Abilities

A good analysis of abilities as solutions to problems raised by rule-following is offered by Julia Tanney (2000). Like many theorists whose work we have been examining, Tanney also draws upon the image of games (in her case, baseball). Games, she says, 'are relevant because they are typically governed by rules', and 'the introduction of rules into the analysis of games seems indicated simply because they provide a standard in virtue of which it makes sense to judge particular moves in the game as "correct" and "incorrect"' (Tanney 2000, 205). When Tanney says 'correct' and 'incorrect' here she is referring to whether a person's actions can be considered to be moves in a game: i.e. given (on Tanney's view) that 'running around the bases' is a

---

[98] We have, of course, not tried to be comprehensive, for there are many other accounts of dispositions we have not considered – an interesting recent example is Yamada 2010, who also says that a proper account of dispositions to follow rules might show that there are a limited set of rules we can be disposed to follow and thus that 'rule following might turn out to be a rarer phenomenon than one might have thought' (Yamada 2010, 293).

form of action made possible by – because it is constituted by – some rule, someone can be said to be playing correctly (e.g. running around the bases) when they comply with the rules for playing the game (i.e. the rule for running around the bases). We have already resisted such talk, but let us continue to see what effect it has on how Tanney pictures the mind learning and working.

The way that Tanney articulates the challenge to be met by her paper is that she is concerned with the conditions under which we can say that someone (whose actions we are observing) is 'making (or attempting) the moves as part of the game of baseball' (Tanney 2000, 205). Here, Tanney rejects an answer that appeals to a 'meta-game', i.e. a game at a higher order where we can be said to be making moves in the ground-level game if we can 'cite the rules of baseball, and to make a case for the material moves being sanctioned by the rules' (Tanney 2000, 205). Tanney is not always clear on how she characterises the 'meta-game': sometimes it is merely a matter of citing the rules (a kind of knowledge report, we may say), and sometimes it is a matter of being 'able to correct or justify [the] moves by appealing to the rules of the game' (Tanney 2000, 206) – in this second case we may interpret 'correct' to include both criticising on the basis of the rules and correcting oneself without engaging in explicit criticism (e.g. trying again, but this time with more attention to the rule). There are important differences here, not only between reciting and evaluating, but also between criticising, correcting and justifying, but let us put these matters aside for they are not essential to the issues we are focusing on.[99]

The dilemma that Tanney sets up is that if we do try to account for how the person we are observing makes moves in baseball (the 'object game') by recourse to the meta-game, then we will need to explain how persons are able to play the meta-game. In fact, because of how Tanney understands games (i.e. as rule-governed domains), she argues that the image of the meta-game is misleading. This is because the ability to follow rules (of the object game) cannot itself be a kind of game (the meta-game), for if it were – and thus if it were itself governed by rules – we would risk triggering the regress. The regress threatens because of the way Tanney understands games, as well as the ideas of correctness and incorrectness (or the

---

[99] This thesis is very much committed to recognising the varieties of normative attitudes, e.g. respect, commitment, justification, criticism, guidance, appreciation, but what is crucial to this thesis is that these attitudes be pictured as tentative, active and creative, and that where any of the above attitudes are described as rule-based (i.e. based on the prior learning or internalisation of rules) that these are not seen to characterise the default operation of the mind.

possibility of error). Thus, if one were to say that the ability to follow the rules of the object-game were rule-governed, and thus themselves require rule-following, then because the moves in the meta-game would need to be performed either correctly or incorrectly,[100] so it follows that there would need to be another set of rules and another game (the meta-meta game), and so on. What Tanney tries to do is stop the threat of the regress at the higher order level, i.e. to describe our ability to follow the rules of the object game in a way that does not require reference to the same kind of 'cognitive-explanatory relation…that we sought earlier between the person who is able to make the moves in the object game of baseball and the rules that govern these moves' (Tanney 2000, 209).

The difficulty for Tanney (and also a difficulty in trying to understand her view) is that she thinks of those abilities (e.g. the ability to follow rules) as themselves somehow 'norms' or 'standards' (Tanney 2000, 209).[101] The tricky thing, then, to hold on to is that the very same thing that is called, by Tanney, a 'norm' or 'standard' (and thus the thing that governs us when we follow the rules of the object game) is also called by her 'an ability.' Examples of these 'other rules' (that are also abilities) are best illustrated by the kind of mistakes Tanney thinks we can make when we follow the rules of the object game: misperceiving, misintuiting, or misreading a symbol expressing a rule; mistaking 'a situation as one that falls (or does not) within the domain of the rule'; and failing 'to act in accordance with the rule, even if one understood it and its domain properly' (Tanney 2000, 216). Because Tanney thinks that mistakes can only be made when they are violations of rules, so she calls these rules, but she also wants to call them abilities (in a way, we have an ability when we avoid breaking these rules).

Tanney dismisses proposals for understanding these abilities (or 'other rules') in terms of innate, implicit or tacit knowledge. The argument against why they are not innate seems to involve an issue of generality, i.e. we might say we have an innate ability to follow the rules of baseball (though it is unclear who would dare to claim such a thing!), but this kind of innateness would not explain why we generally have abilities to follow rules (including, but not limited to, the rules of baseball; Tanney

---

[100] For Tanney, at least on the incorrectness side, it is a 'platitude that error is diagnosable only in so far as a norm or standard has been violated' (Tanney 2000, 206). This 'platitude' was challenged in Chapter 6 above.

[101] Indeed, Tanney is very free with her terminology, sometimes referring to 'norms' and 'standards' as opposed to 'rules', but sometimes also calling those 'norms' or 'standards' 'other rules' (see, e.g. Tanney 2000, 210).

2000, 210-1). The argument against tacit and implicit knowledge[102] is that it does not offer a solution to the problem, saying only that the 'vehicle of knowledge' can differ in terms of whether it is not carried at all or potentially carried by consciousness. Tanney argues that the same regress problem arises whether one makes the vehicle of knowledge here conscious or not, personal or sub-personal (Tanney 2000, 213).

Innate, tacit and implicit knowledge, then, do not help. Tanney is after an explanation (of abilities to follow the object level rules) that is general enough, and one that offers a solution (i.e. makes sense of how the person we are observing can go wrong with respect to the those object level rules) while not triggering a regress. Innateness is not general enough, and tacit and implicit knowledge does not get at the issue (it is still subject to the sting of the regress).

But what does offer a solution? In fact, Tanney does not so much offer a solution, as two retreats: first she says that sometimes we may be satisfied that the person we are observing is acting in accordance with the rules (without needing to know more), but sometimes, if we want to 'rule-out mere accidental conformity', we may try to inquire further (e.g. ask her to cite the rules – for what else?); and second, that perhaps the very abilities (or 'other rules') that we have been trying to investigate are 'part of the "bedrock"' (Tanney 2000, 224).

There are several things we can observe about these conclusions, and in general about Tanney's paper. The first is to question the very pursuit (which she ends up pegging back somewhat) of trying to know for sure – trying to guarantee – that the person we are observing is really, truly, playing the game. Tanney pegs this back a little by saying that it may depend, and that sometimes we may be satisfied with observations of 'mere conformity', but she must realise that when she advocates inquiring further (and asking the person to cite the rules) that she limits the scope of the inquiry (concerning rule-following) in two ways: first, it limits it to those persons we can ask that of (and thus excludes pre-linguistic infants); and second, it limits what can be reported (for the theorist would be looking for representable rules, the point being that this would restrict the inquiry to explicit rules).[103]

---

[102] The difference, according to Tanney, is that whereas 'implicit knowledge is available to consciousness once it has been made explicit', 'tacit knowledge is not normally accessible to consciousness at all' (Tanney 2000, 211, fn. 12).

[103] Of course, one would also need to add a proviso as to the questionable credibility of such 'reports' (see Section 10.D.).

From the perspective of this thesis, this limiting of the inquiry – and thus this limiting of the explanatory appeal to rules and rule-following – is a welcome one. In terms of Tanney's conclusions, it illustrates two things: first, that Tanney is right that we need not know for sure when someone is really, truly, playing 'the game' (i.e. where 'the game' is understood by us to be identifiable by certain criteria, and thus exhausted by certain rules), but that what this means is not that there is a fact of the matter about whether someone is following the rules, but simply that we need not always use rules to decide, for practical purposes, whether we will treat someone as playing a game;[104] and second, that there is no 'bedrock' – no ultimate, invisible, 'other rules' (or abilities, if we define abilities in Tanney's sense of it), that we reveal when we follow or violate allegedly more particular, object level rules.

There is no 'bedrock' because it is not the case that we are always following the object level rules; one only needs the bedrock (granting some special status to certain 'other rules') when one wants to establish that we are always following object level rules when we are doing something (e.g. always following the rules of chess when playing chess) – but as soon as we see that we need not always be following the object level rules to be playing chess (because playing chess is not necessarily and / or sufficiently a matter of following rules), but rather that there may be some limited, non-exhaustive rules (which we may call 'the rules of chess') that we can use to teach others to play chess or criticise them (this thereby being just one way in which can teach them chess or criticise their chess playing), then we also see that we do not need some special, higher-order or more general rules (or abilities) that, as it were, secure that we are always following the allegedly object level rules (even when we do not realise it).

The general point is this: talk of abilities should be not subject to talk of following rules. Abilities are themselves not rule-structured, i.e. not a matter of following rules, and nor are abilities always abilities to act in accordance with rules. One suspects, from the way that Tanney sometimes does talk of abilities, that she wants to release abilities from having a rule-like form or having rules as their object, but the problem is that Tanney begins with an understanding of games as exhaustively rule-governed domains, with the result that any epistemology or account of cognition

---

[104] Another way to put this is that games are not the sort of thing that are exhaustively rule-governed, and that is why we cannot rely exclusively on the rules to tell when someone is playing the game.

that she goes on to consider is one that she thinks must be describable under the auspices of the rules that she thinks govern that domain.

Our point in this Part of the thesis is not to show that games (or other activities) are not exhaustively rule-governed domains (that was the task of Part II), but rather that it is implausible to try to account for the mind learning and working as if it was always, by default, following rules. What we have seen in Tanney's paper is that she has tried to give us an account of rule-following via the concept of abilities, but that she could not give us an answer as to how these abilities worked (she could only point to how those abilities could not work) – our point being that this is understandable, for abilities cannot be described under the watchful eye of rules. This is vital, for, as we shall see, persons participating in social life – when this understood from the second-person perspective – do so, at least in part, thanks to abilities that need not be formed (learned, developed) on the basis of pre-existing rules, nor take rules as their object.

Somewhat ironically, then, one of Tanney's conclusions is a welcome one, i.e. any inquiring into rule-following (whether someone was really following rules) is going to be a very limited one – not only because we will not always undertake it or need to undertake it, but also because it is restricted to asking those who can give reports, where these reports need to contain explicit rules, though with the proviso (in keeping with what was said in the previous Section) that any such reports cannot be taken on face value. Tanney's other conclusion needs to be reversed: these limitations show not there is a 'bedrock', but precisely that there is not one.

F. Blind, Unconscious and Implicit Rule-Following

We are now approaching the conclusion of our discussion of varieties of rule-following, but we still have one more option to explore, i.e. those accounts that appeal to blind, unconscious or implicit rule-following (to put these all in one basket is not to assert that there are not theorists who use them in specific and distinguishable ways, but it is to suggest that the general strategy is similar). The key issue for theorists who speak of blind, unconscious or implicit rule following is to what extent – if necessarily at all – our being guided by a rule is mediated by some conscious cognitive process. The problem that arises is whether, if we say that it can be and / or often is not so mediated, we can still refer to that process as one of being guided by a rule (of course, one interesting question in the background is why we should want to

characterise such unconscious processes as being guided by a rule). The dilemma is well illustrated in José Zalabardo's paper, 'One Strand in the Rule-Following Considerations' (2009).

Zalabardo asks: 'What happens when I decide to apply the word "cube" to a given object, or to write a certain number as the next term in a number series?' (Zalabardo 2009, 509 – here it must be said that the second example is the better one, for 'deciding to apply' something already sounds like a very conscious, deliberative process). I can do so 'mechanically, without the intervention of conscious processes', or I can do so in a way where 'I see myself as selecting my response under the guidance of an item that is present to the mind and intimates to me how to respond in this case' (Zalabardo 2009, 509). Zalabardo's strategy in the paper is to reject a dualism between doing something mechanically and doing something under the guidance of something, at least where such a dualism has the effect that we think of guidance as necessarily mediated by a conscious process. Instead, he splits the idea of guidance into two forms: first, a 'literal' sense of guidance; and second, a 'symbolic' sense, and he argues that 'in the literal sense the guidance story is always false, but that it in the symbolic sense it may be true of any of the episodes that we are inclined to describe in these terms' (Zalabardo 2009, 510).

According to Zalabardo, the literal sense falls prey to the usual problem of regress. Return to the cube example. The question is how the cube is present to the mind. Is it present to the mind because I relate it to a picture of a cube? And if I do so relate it, do I do so only insofar as I interpret that picture as one that can be applied to what I am seeing? If I answer 'yes' to these questions, I am subscribing to the literal sense of guidance, and falling prey to the regress of interpretations problem. Zalabardo wants to reject (and he also thinks Wittgenstein argued that we ought to reject) the literal sense. But now the crucial move that interests us is how Zalabardo articulates the alternative (this alternative is what he calls the symbolic sense of guidance). A number of different formulations appear in the paper. One is that 'cubicalness could be present to the mind directly, not only by being suitably related to the picture of a cube that is present to the mind' (Zalabardo 2009, 512), though no more is said here about what it mean for it to be so present. One of the interesting things to observe about this formulation is that it does not say that a cube could be present to the mind directly, but rather 'cubicalness', and it is useful to keep in the

background the question why it should be cubicalness and not cubes that Zalabardo wants us to see directly.[105]

At another moment, Zalabardo (allegedly following Wittgenstein) says that the process must not be one where I feel I am waiting for instructions or waiting to be told how to obey a rule; rather, my response must 'feel spontaneous' – whatever it is that is presented to us is done so in a way where I immediately know how to go on, and that is how I 'feel inclined to proceed' (Zalabardo 2009, 514). Here we have two different candidates: spontaneity and inclination, or perhaps one with spontaneity better read as an adjective, qualifying the noun (inclination), so that what we are after is a spontaneous inclination, or an inclination felt to be spontaneous (notice that even here, the notions of 'feeling', e.g. that I am not waiting etc, tend to point to some level of awareness of what I am doing). But here we can ask the same question: why think that the 'spontaneous inclination' must be an inclination to follow a rule? Why is this 'spontaneous inclination' a way of articulating how we are guided by something?

We receive a clue about what is going on here when Zalabardo draws on Wittgenstein's dismissal of intuition as an account of guidance. Wittgenstein asks two questions about intuition: 'If intuition is an inner voice – how do I know how I am to obey it? And how do I know that it doesn't mislead me?' (Wittgenstein 1953, §213; quoted in Zalabardo 2009, 513). Zalabardo fastens on to this second question: we do not want our account of guidance – of how I am guided – to be 'normatively inert' (Zalabardo 2009, 514); it must somehow inform me about how I ought to go on, but (according to Zalabardo) it cannot inform me about I how I ought to go on in a way that would make me second-guess myself – to ask myself that first question (i.e. but is this really what I ought to do, is this rule really what I ought to obey?) – for if I did second-guess myself, then I would enter the regress of interpretations.[106]

---

[105] The one explicit thing that Zalabardo says about this is that 'cubicalness…will be able to guide my uses of the word "cube"' (Zalabardo 2009, 513), but this is just to rephrase things – the question we need to keep asking is why it is assumed that there is a need for saying that my use of a word needs to be 'guided' by something.

[106] From the perspective of this thesis, a lot depends on how extreme one pictures second-guessing. If one pictures it as necessarily leading to paralysis, where one cannot do anything and cannot trust oneself at all, then this is understandably characterised by Zalabardo as negative. But we can be more subtle here about this process: it can be depicted as something more tentative, a kind of waiting and suspending of action that need not be, or lead to, paralysis. In that more subtle sense, 'second-guessing' is not a negative thing that must be ironed out: it is part of the tentativeness of normative attitudes, and is one of their most important features (in other words, the very advantage of the normative is that it creates this room – room, one might say, for the creation of reasons).

Zalabardo calls on Crispin Wright's interpretations of Wittgenstein to assist with the formulation he is after. Given what has been said above, the kind of formulation Zalabardo wants is one that is not normatively inert, but one also that is not a matter of second-guessing, e.g. of the training 'feel[ing] inclined to respond in certain ways to new cases, with what Wright calls "a sense of no option"' (Wright 1980, 222; quoted in Zalabardo 2009, 516).[107]

Zalabardo's paper points us in the direction of Wright – in a way, the paper attempts a vindication of the direction in which Wright takes Wittgenstein – and so it will pay to look briefly at Wright. But before we do so, we can make a number of observations about the above analysis by Zalabardo. Let us consider the following array of possible experiences: first, I see a cube (not cubicalness) directly (this is not a category discussed by Zalabardo); second, I see cubicalness directly (this is what Zalabardo wants); and third, I see cubicalness indirectly (this is what Zalabardo suggests is impossible, or rather falls into the regress argument).

Zalabardo does not recognise the first possibility, and the reason for this seems to be that he wants to describe the incident of us seeing a cube as one in which we might be warranted to see one. The entire discussion is set up as one in which I am always prepared to answer how it is that I know I am really seeing a cube. It is because Zalabardo places this question perpetually in the background that he tries to make work a description of the mind's encounter with cubes as a matter of directly seeing cubicalness, of being guided in a way by something according to which I can guarantee access to a cube, and where a mistake (a misperception, say) can be made sense of on the basis that I somehow failed to follow the rule (failed to apply cubicalness to the cube). The question to ask back is: why this question perpetually in the background? Why the implicit pressure on persons to be able to answer it?[108]

It is important to see what happens when we loosen the hold of that question, i.e. we immediately see that it is possible to speak of appreciating a cube in front of me, though not in a way in which I apply some general criteria (implicitly), but in a way that may create a way in which one can appreciate one is encountering a cube (though this may be risky, because others may not agree that we appreciate as a cube

---

[107] Notice that it is not just a matter of responding, but responding with some sense of there being no other way.

[108] The point here is not that we cannot sometimes feel the pressure of this question: clearly we can. The point is not to posit it perpetually in the background, for this has deleterious effects on how we picture the mind learning and working.

ought to be counted as appreciating cubes). The fear of doubt, impregnated by the above question positioned perpetually in the background, leads to a picture of the mind that can never take a risk, and in not being able to take a risk, is always only applying general criteria to what it sees. As a result, the picture of the mind becomes passive, rigid and static: an obedient servant to an always and already stable and secure ontology (of cubes as cubes according to such-and-such criteria).

A third possibility was mentioned above, namely: seeing cubicalness indirectly (which Zalabardo says is impossible, and which he asserts was the target of Wittgenstein's criticism). Let us see, given what we have just said, whether we can save this possibility. Here is Wittgenstein cited by Zalabardo:

> What really comes before our mind when we understand a word? – Isn't something like a picture? Can't it be a picture?
>
> Well, suppose that a picture does come before your mind when you here the word 'cube', say the drawing of a cube. In what sense can this picture fit or fail to fit a use of the word 'cube'? – Perhaps you say: 'It's quite simple; - if that picture occurs to me and I point to a triangular prism for instance, and say it is a cube, then this use of the word doesn't fit the picture.' – But doesn't it fit? I have purposely so chosen the example that it is quite easy to imagine a method of projecting according to which the picture does not fit after all.
>
> The picture of the cube did indeed suggest a certain use to us, but it was possible for me to use it differently.
> (Wittgenstein 1953, §139, quoted in Zalabardo 2009, at 510)

Zalabardo reads this as an argument against the literal sense of guidance – that it could not be a picture I relate the cube to, for this would require me to interpret the picture, and would thereby trigger the regress. But what if we read this passage as one in which Wittgenstein is pointing to the ability of the mind – and the experience we can have – of relating tentatively with cubicalness, i.e. with the qualities we associate with cubes? Part of this tentative relation with cubicalness may be wondering about or questioning some set of rules or even pictures, but having such a tentative relation is not a matter of trying to apply, but then noticing some obstacle to applying, some default rules. In other words: we first have an experience (indirectly, of cubicalness), and as a result of this experience we may come to challenge any criteria we use to articulate when something counts as, e.g. a cube. Imagine we see what seems to us to

be a pyramid, but turns out to be better appreciated as a cube[109] – it need not be the case that we must have been guided by what counts as a cube,[110] but it can be the case that we might then go on to say that one novel way of appreciating cubes is to approach them, in such-and-such circumstances, as if they were pyramids (thereby creating a new association, which we might go on to articulate as a rule). In other words, we might, as a result of appreciating a cube in this new way, add something to our experience of cubes – the experience 'suggests a certain use to us', and we come to see we can use the word cube 'differently.'

We do not wish to enter Wittgensteinian exegesis here – that is not the key. Aside from what Wittgenstein might have meant, we can here introduce a kind of normative experience where we wonder about and hesitate over what something ought to be: ought this experience be part of our understanding of cubes, or perhaps pyramids (or is it an irrelevant experience altogether)? And – we might ask ourselves – what does this experience say about any of the rules or pictures we might have previously used to articulate what counts as a cube? We need, then, to see a gap between any way in which might articulate, on occasion, what a cube is (implying that we ought to see, in the future, cubes when they have such-and-such features), and the exercise of appreciation: no way of articulating those features maps onto, or controls (though it can contingently guide and be used as a standard of criticism) that exercise.

How does all this feature in Wright's work? Recall that Zalabardo's point was that Wright shows us how the symbolic sense of guidance needs to work and can work. We can tell how it might work already in the first sentence of one of Wright's papers – the most pertinent one for our discussion here (Wright 2007): 'It is natural to think that in any area of human activity where there is a difference between *correct* and *incorrect* practice, [what] we achieve is (partly) determined by rules which fix what correct practice consists in, and which in some manner guide our aim' (Wright 2007, 481; original emphasis – the 'partly' here is indeed at issue, but Wright does not elaborate on what he means by that 'partly').

---

[109] It is important to see that this is not determined – we could just as easily have decided to treat this experience as relating, indirectly, to pyramidness.
[110] Though it is also possible for us to have had in mind cubes and criteria for something counting as a cube (imagine we have just been to a 'cube' convention) – this is possible, but only implausibly necessary.

The way Wright sets up the dilemma, then, is like Zalabardo's: the question is when a person can be said to have followed a rule or not, and we need this question because we need to be able to say when we have really made a move in a game, or seen a cube, or whatever action in whatever 'activity' it might be that we have defined as being governed by rules.[111] Once this is in place, the entire picture of the mind learning and working is subject to the question: '*How* does a rule actually manage to lead us?' (Wright 2007, 482). It is this subjection, of course, that we are resisting in this Part of the thesis. Where does it take Wright, i.e. how does Wright picture the mind by fastening on this question?

Wright argues, as his title suggests, for a conception of 'rule-following without reasons.' We need this because it is untenable to think that we need to have something 'in mind' whenever we follow rules; or that our every action of following a rule needs to be 'mediated by the internalisation of explanation-transcendent instructions' (Wright 2007, 487). Here is one of the clearest statements of the position Wright takes:

> Basic cases – where rule-following is blind – are cases where rule-following is *uninformed by anterior reason-giving judgement*… In such a case one follows a rule 'without reasons' in the precise sense that one's judgements about the input condition for correct application of the rule are not informed by the exercise of concepts other than that which the rule concerns – that is, the concept whose expression the rule regulates and grasp of which consists in competence with that very expression. Such a judgement is an ungrounded response in the precise sense that is not to be rationalised by the modus ponens (as I have suggested, the only possible) model – by the picture of rule and input as (potential) components of independent thought. It is still essentially the response of a rational subject, and still be appraised within the *categories* of rationality – justification and truth. But it is an action for which, it is now tempting to say, the subject has and can have no reason – for the possession of such reasons and their appreciation as such would demand the exercise of an anterior concept, in an independent judgement, of what made the action appropriate. (Wright 2007, 497; original emphasis)

Here, Wright comes very close to Tanney's position: the rules just are the abilities, and because we can think of abilities as not necessarily 'mediated by the internalisation of explanation-transcendent instructions', so we can make sense of the

---

[111] As Wright puts it, e.g., 'a judgement expressing a putative such recognition' (of an object, say) 'insofar as it can be correct or incorrect, must presumably be a rule-governed response' (Wright 2007, 484).

notion of following rules blindly.[112] But Wright makes the point in a slightly different way, for he suggests that perhaps when we speak (somewhat misleadingly) of following rules blindly, we ought not to speak of following anything at all – perhaps, he says, we 'are not really guided by…anything' (Wright 2007, 497). Then he adds:

> The problematic invited us to try to construct an account of what, when we follow a particular rule, constitutes the facts about the direction in which, step by step, it guides us and how we are able to be responsive to its guidance. But in basic cases the invitation emerges…as utterly misconceived; for it presupposes a false conception of the sense in which basic rule-following is rational. Basic rule-following, like all rule-following, is rational in the sense that it involves intentionality and a willingness to accept correction in light of error. But that is not to say that it involves responsiveness to the requirements of the rule, conceived as instructions, as it were, which can feature in thought and rationally inform one's response. (Wright 2007, 498)

At first blush, this is promising from the perspective of accounting for how the mind works and learns without having to argue how it is (always / necessarily) guided by rules (or by any other normative resources). In fact, however, it digs us deeper into the rule-following picture. To see this one needs to see that when Wright speaks of responsiveness as a kind of willingness to be corrected (rather than a responsiveness to the rules conceived of as instructions), he is still envisaging being corrected on the basis of the rules. Recall the opening sentence of Wright's paper: what is correct and incorrect is determined by the rules. But, for Wright, it is somewhat misleading to speak of 'following rules', for it suggests that the rules are somehow anterior to or prior, or exist in some autonomous realm, outside of us, and guide us as if we were puppets. Rather, the rules consist in competences – we should not speak of being guided, for it makes no sense (it adds an unnecessary object of explanation and only creates problems) to say we are 'guided by' competences. The case that Wright has in front of him is a case where the competence in question is one that is already identified by a rule – what he wants to do is avoid the problems that arise when we think that the competence is separate to the rule, such that the rule is prior to the competence and guides it. For Wright, we can act rationally, for we can act competently (i.e. exercise our competences), and what makes it the case that we can

---

[112] Indeed, towards the end of the paper Wright speaks of 'competences' (Wright 2007, 498), which he seems to use in the same way Tanney would use, i.e. they do not presuppose rules, for they are the rules (the 'other rules' in Tanney's terminology).

act rationally insofar as we exercise our competences is that our competences just are the rules.

We can see that Zalabardo's reading of Wright is not quite accurate, for he fails to speak of how Wright says that it is misleading to speak of being guided by a rule or following a rule – misleading insofar as it makes us think of the rules as transcendent instructions somewhere anterior to us (and thus leads to a kind of Platonism about rules). The question of consciousness is not quite the problem for Wright as it is for Zalabardo, though perhaps what is more accurate to say is that what matters to Wright is how we understand consciousness in this context (i.e. whether we understand it as a matter of being guided by some rules-as-transcendent-instructions).

What is vital for present purposes is seeing why Wright digs us deeper into the rule-following picture, and thus why he makes us miss all that we miss when we picture how the mind learns and works exclusively on the basis of following rules (or exercising competences that are, in fact, rules). Above it was said that when Wright speaks of responsiveness as the willingness to be corrected, he is still thinking of us being corrected on the basis of the rules. What Wright misses, then, is precisely the way in which we can be responsive, and thus willing to be corrected, in a way that reveals the limitations of the rules. This way of thinking about responsiveness is not open to Wright because his starting point is correctness/incorrectness-determined-by-rules. Being willing to be corrected is inconceivable in any other way than on the basis of the rules.[113] The point here is not that making room for being responsive means we cannot be responsive to rules; we can be responsive to rules, but we need not be responsive to rules all the time – responsiveness to rules is just one way in which we can be responsive. But to see that we can be responsive in a way that is not dependent on being responsive to rules we must allow for something tentative and risky: for creating a way in which one can experience something, e.g. a cube (this being tentative because we ourselves might not endorse it, and being risky because others might criticise us for it). Making room for this kind of tentative and risky responsiveness also helps us to see that our treating of some criteria as relevant is an active treating: it is not the default mechanism; if it was, we would be describable as slavish, passive, rigid, static, obedient rule-followers.

---

[113] His 'partly determined' suggests there may be room for something else, but we cannot assess this without knowing more about he means.

Consider the following case: imagine that I have only ever seen cubes as pictures on a page. What I consider to be a cube – let us imagine I have thought about it and tried to articulate it – is that it appears in the way that cubes as pictures on a page have always appeared to me. Then imagine that I see something that appears to me in a way in which it might be a cube or it might not (the matter is not foretold). Nothing determines whether I see this as a cube, or I see it as not a cube, i.e. I do not mechanically, spontaneously, immediately, etc apply a rule and categorise the object before me. Instead, I may challenge my previous articulation of cubes, and now refer to cubes as cube-pictures-on-a-page and to 'shubes' – these being, say, big marble blocks. Wright would set up this case in such a way where there is a fact of the matter as to whether a marble-block can correctly be called a cube. When I see that object, I either see it as a cube or I do not see it as a cube (at which point I see it as something else). When I see it as a cube, I exercise a competence (in effect, a rule) to see it as a cube. I therefore recognise it as a cube correctly or not, for it really is a cube or it really is not a cube (for the rules – which anyone, presumably, in the relevant community, can access – dictate whether it is a cube). It is a very neat explanation, but the neatness is achieved at the cost of positioning us passively with a world already categorised in a certain way, and now awaiting correct recognition of that categorisation. What we need to make room for here is the experience of the environment as indeterminate – as not being in a certain way, such that this enables us to explore (tentatively, and in a way that is somewhat risky) how we might want to understand it. In the course of doing that, any previous articulation may become not so much inapplicable (in the sense in which I recognise that what I see is not within the scope of the rule), but simply irrelevant – and potentially (but not necessarily) superseded by an articulation that changes any particular prior articulated grouping. Allowing for this tentative and risky relation with the environment helps us avoid a picture where the mind learns to categorise and then just re-enacts this categorisation.

There is a very interesting moment in one of the above passages by Wright (which was cut out above, so that we would not be distracted by it at the time). Immediately after saying that the basic cases of rule-following (blind rule-following) are those where rule-following is uninformed by anterior reason-giving judgement, Wright calls upon an analogy and says that these cases are 'just like the attempts of a blind man to navigate in a strange environment' (Wright 2007, 496). Then comes a very instructive footnote attached to this analogy: 'Of course, the analogy limps

immediately after this point – the movements of the blind man will naturally be hesitant. But in basic rule-following "I act quickly, with perfect certainty, and the lack of reasons does not trouble me"' (Wright 2007, 496, fn. 19 – the internal quote is Wittgenstein 1953, §212). Now this is a fascinating aside from the perspective of this thesis: combine it, for a moment, with the idea of rules only 'partly' determining correctness. We get the possibility – only hinted at by Wright – that we can have experiences of strangeness, of hesitancy, and in those experiences the rules (or competencies conceived of as rules) are not really at work, or not dominantly or exhaustively at work. A way of reading what this Part of the thesis is about is to take seriously the blind man in a strange environment – the point being that we miss him, and the way that the analogy limps, when we focus exclusively on the question of how we do, in fact, follow rules.[114]

It was mentioned earlier that the last category to be considered in this Chapter was 'blind, unconscious and implicit rule-following'. We have not yet referred to 'implicit rule-following', and so must do this as a final task. A useful source for this task is chapter 7 (entitled 'Regularities, Rules, Meanings, Truth Conditions and Epistemic Norms') of Paul Horwich's recent book (2010).

Horwich's argument, which focuses on the sense in which language is rule-governed, is, in summary, that: 'the basic facts…are law-like regularities of word use (characterised in non-semantic, non-normative terms); that such regularities help engender (i.e. are the primary reductive basis of) facts about which rules of use we are implicitly following; that these facts suffice to fix what we mean by our words and hence sentences; and that the meanings of our sentences (given contextual factors) determine their truth conditions – which we ought to desire to be the conditions in which they are accepted' (Horwich 2010, 113-4). We will not here engage in evaluating this argument in total – what we are interested in is how Horwich presents the need for an account of implicit rule-following, and what how he characterises that kind of following of rules.

Part of the job of explaining implicit rule-following is taken up by characterising what implicit rules are. This is undertaken by contrasting explicit with implicit rules. Explicit rules are like the rules of games such as chess and football, or

---

[114] Of course, to refer to a 'blind man' in a 'strange environment' is already to suggest something marginal and negative: our point is to say that what the image referred to by Wright suggests is something central and positive.

the rules of a club, the rules of parliamentary procedure, the laws of certain countries, and even 'the rule that I impose on myself for how many times a week to go to the gym' (Horwich 2010, 113). What binds all these rules together is that they are 'explicit instructions' that we typically 'hear or read…or think…to ourselves'; 'we understand these formulations – that is, we attach semantic contents to them; we then decide to do what is dictated by those contents; and so we conform more or less successfully' (Horwich 2010, 113). Implicit rules, by contrast, are not formulated. Horwich's examples are the 'rules of English grammar' and the rules that determine 'how far apart two people stand from one another when they are having a conversation' (Horwich 2010, 114). It might appear surprising to include the rules of English grammar, for presumably Horwich would agree that some of these are written down. But here is where, having defined these rules in contrast to explicit rules, Horwich shifts the explanatory focus not on the kinds of rules but on the kind of rule-following they engender, i.e. implicit rule-following. Again, he draws on a contrast: explicit rule-following involves the following steps: 'the rule is spelled out, and that formulation, in virtue of its meaning, is allowed to guide our activity'; implicit rule-following, however, 'is more of a theoretical posit and more murky', but the one thing that can be said about it, at least initially, is that 'it's not the same as unconscious rule-following – because that could be a matter of unknowingly, but explicitly, following a rule that is formulated in one's language of thought' (Horwich 2010, 114). We shall here put aside Horwich's definitions of explicit rule-following, including 'unknowing explicit-rule following' (which reads somewhat mysteriously), for our focus is the implicit kind. It is this kind, says Horwich, that is really at the heart of philosophical problems to do with rule-following. Here is how Horwich defines it:

> Person S implicitly follows rule R! if and only if:
>
> > (a) S's activity is governed by the ideal law R
> > (b) There is some tendency to correct instances of non-conformity (i.e. to react against his initial inclinations)
>
> Here, the role of condition (a) is to specify which aspect of the agent's activity fixes the content of his rule – i.e. what makes it R!, rather than R*!, that is being followed. And the role of condition (b) is to explain why S may be regarded as following a rule, rather than as merely obeying a natural law (which is what, for example, the planets do in orbiting the sun). For a tendency

to self-correct manifests sentiments of relative dissatisfaction and satisfaction, and thereby encourages talk of 'goals', of 'what ought to be done', and of 'violations'. (Horwich 2010, 117-8)

Clearly, we need to know more about what is an 'ideal law' and how it governs S's activity, and we also need to know more about the claim that self-correction reveals rule-following.

Horwich's notion of an ideal law is quite peculiar: it is not a rule in the sense in which we explored that concept in Part I and the sense in which the theorists we have been considering so far have used it, but rather more like a natural law, though one with 'exceptions' (Horwich 2010, 118). By 'exceptions', Horwich means that built into the law is recognition of 'a variety of factors that cause deviations from the behaviour that would "ideally occur"' (Horwich 2010, 119), e.g. 'only in certain "ideal" conditions do the orbits of the planets obey Kepler's laws' (Horwich 2010, 119). This ideal law, then, is markedly different to a rule, about which it is usually said that it can be disobeyed precisely in contradistinction to natural laws that are often said to be un-breakable. It has to be said that it is not easy to make sense of the way in which the conditions Horwich refers to are 'exceptions'; rather, they read more like (and are usually explained in philosophy of science) as being conditions under which (the results of the tests we have conducted so far show that) the law is not applicable as a description, i.e. circumstances in which it is not a valid description of the behaviour of objects being analysed. The point is important because it raises the question of in what sense person S can be said to 'notice instances of non-compliance' with the ideal law. Horwich seems to want to present what are effectively conditions of applicability as conditions under which the law is violated. He needs to do this because this is what he means by non-compliance (which is an element of condition (b)). But if what Horwich presents as 'exceptions' are better understood as conditions of applicability, then non-compliance (in the sense of a violation) could never be noticed, for it would not be non-compliance that would be in play when the law is not valid – it would just be that the law is not relevant to the description of the behaviour of objects in those circumstances.

There are two further problems with the notion of an ideal law, both of which Horwich recognises, but only one of which he thinks is a potential problem: first, Horwich qualifies his drawing on 'ideal laws' by pointing out that we can only refer to them as 'scientifically best idealisations', meaning that given that there may

'sometimes be more than one "equally best" idealisation', there may be 'a certain degree of indeterminacy as to which rule is being implicitly followed' (on the condition that condition (b) is satisfied; Horwich 2010, 118, fn. 7); second, Horwich recognises, but does not think that this is fatal, the possibility (some might put it stronger, the inevitability) that we have not (and possibly could never) list all of the various possible conditions under which the ideal law would not be valid, i.e. that there are a potentially infinite number of 'exceptions' (to use Horwich's terminology). Horwich thinks the first of these is a potential problem, but just bites the bullet. In terms of the second, he does not think it is fatal because it is only fatal if one assumes that one must describe the ideal conditions *a priori*; instead, says Horwich, it is *a posteriori* matter (Horwich 2010, 120). This is perfectly reasonable as a thesis about how we come, in a never-exhaustible way, to refine / revise ideal laws – but it is not clear how such an *a posteriori* description is compatible with the claim that we can identify 'the' implicit rule that we are following.

There are difficulties, also, with condition (b). In fact, Horwich himself qualifies it early on in his paper by saying (in a footnote) that 'We can't simply identify cases in which S deviates from his rule with cases in which he corrects himself (or has a disposition to do so) – because he may well fail to notice certain cases of non-compliance' (Horwich 2010, fn. 6, 118). For this reason, Horwich says, 'S's rule cannot be straightforwardly read off his practice of self-correction', but he thinks that 'nevertheless, that practice is an important part of the empirical evidence that can help us…to reach plausible conclusions as to which combination of ideal laws and occasional distorting factors are influencing S's activity' (Horwich 2010, fn. 6, 118). Above, when discussing Stueber's account (in Section 10.D.), we already gave many counter-examples against the idea that the tendency to self-correct can be taken as evidence (even if qualified as 'plausible') of rule-following, so we shall not repeat this again now. Here, we simply make note of both Horwich's qualifications and those counter-examples.[115]

But these are not the only problems with condition (b). For one, it is difficult to see what kind of practice of self-correction could tell us (the observer) that an ideal law – and that particular one – was being followed. Part of the difficulty is that, by

---

[115] Though it is important to see that none of the counter-examples offered above were based on the potential problem identified by Horwich, for Horwich's problem depends on assuming non-compliance as a matter of fact to be noticed, rather than as an evaluation to be made.

definition, the practice of self-correction does not show that a rule was being followed because we criticise ourselves for *not* following the rule; here to maintain a link between self-correction and ideal-law-following one would have to say that our self-correction shows us that we were attempting to follow the ideal law (as we usually do successfully), but then 'attempting to' surely suggests some kind of understanding of the rule which presumably Horwich would not want (perhaps more problematically, would we not need evidence to rule out the possibility that it is only for the purposes of self-correction that we refer to some ideal law?). But even granting some way round this, we would need to show how some practice of self-correction – Horwich refers to 'occasional self-correction manifest[ing] selective dissatisfaction, and hence orientation towards a goal' (Horwich 2010, 123) – indicates that it is that particular ideal law that was (attempting to be) followed. Is Horwich suggesting that some kind of grimace would tell us which law is being followed? Or is Horwich requiring the person to have some understanding of the ideal law and to verbalise this in the moment of self-correction? It is difficult here to avoid the conclusion that the cases Horwich is envisioning are cases where he assumes that there is such-and-such an ideal law governing S's activity, and then simply suggests that we interpret S's (occasional) self-correction (grimaces?) as signs that ideal law was the one that S was committed to / attempting to fulfil, and presumably the one that S on the whole follows etc.

The difficulties with Horwich's account revolve around the relation between his condition (a) and condition (b). It is in order to construct a bridge between them – between ideal laws and practices of self-correction – that Horwich characterises ideal laws in certain ways (i.e. as subject to exceptions, rather than applicability conditions) and also introduces ideal laws as the basis of self-correction. Horwich seems to want a perfect fit between them: we follow certain ideal laws, and we can know that we follow them because of the way we criticise ourselves on the basis that we did not comply with them. This looks fine from a distance, but the moment one notices that ideal laws are not the kind of laws that cannot be violated, and one notices that self-correction practices do not in any way point to (unless we assume they do) some particular ideal law as a basis for correction, is also the moment in which the entire device of 'implicit rule-following' collapses. The very implausibility of Horwich's definition and use of ideal laws should at least make plausible the conclusion that there are no such things as 'implicit rules', and no such thing as 'implicit rule

following' (at least as he defines it). Notice, though, that none of this criticism of Horwich's account ought to make us think that it is impossible to suggest that there may be ways in which we can learn explicit rules, such that they help us form certain abilities; of course, whether the abilities that are formed are best described (or, worse, only describable) as an unconscious / sub-personal processing of explicit rules is another matter (indeed, it has been argued above we should resist this reading). This is not, however, what we get in Horwich – what we get with his account is the implicit following of implicit rules (and this, we have argued above, is incoherent).

We have not covered here the full range of the varieties of rule-following, but we have engaged with it in quite some detail, at least sufficient hopefully to give us a sense of what is being missed when it is proposed as a general theory of how the mind learns and works. We have proceeded slowly and carefully, for many of the issues that have arisen above offer a good overview and introduction to the matters raised in general by the second element, the governed minds element, of the governance view. This also means that the Chapters that follow will be shorter and more focused on specific contexts in which some of the general issues have arisen. The first of these contexts, considered in the next Chapter, is the debate over the scope, limits and viability of the notion of computation.

To recap: this thesis does not argue against the possibility, or even the more-than-occasional recurrence, of rule-following. It simply argues against a depiction of the mind that represents the mind as a rule-following thing – as something that necessarily follows rules, or as being necessarily governed by rules. Quite often, as we have seen, such claims seem reasonable because of the way that the examples are set up: that which can be done, and ought to be done, is itself presented as demarcated by rules: the mind, then, either does the correct or appropriate thing (does what the rule requires) or does the incorrect or inappropriate thing (violates the rule, and in that sense malfunctions). This has the effect of squeezing out the kind of tentativeness, uncommitted exploration, capacity for surprise, flexible learning, and inherent revisability, which, amongst other things, are so vital from the second-person perspective. In doing so, it risks narrowing our understanding of the social, and furthermore neglecting to see and describe the fuller array of abilities we need to participate in social life.

Chapter 11. Computation

The scope and the very viability of the notion of computation have received extensive attention in the philosophy of cognitive science, as well as in contemporary linguistics. As with the immediately preceding Chapter, the aim is not to give a summary or overview of this literature, but to engage with the philosophical arguments and examples with a view to evaluating what kind of picture of how the mind learns and works they bestow on us.

A computational account of how the mind learns and works has been said to have emerged from two interrelated streams: first, the idea that mental life can be construed in terms of propositional attitudes (and inferences from them or between them); and second, the idea that language processing involves rules operating on language-like representations (Abrahamsen and Bechtel 2006, 159). Both of these streams take it that the systematicity of the mind is an essential feature of it (for some, it is even the explanatory target of our understanding of mind). The advantage of proposing propositional attitudes and analysing the viability of inferences from and between them, or studying the manipulation of rules in symbolic representations, is that one can construct a coherent system – a symbolic or representational architecture, as it is sometimes called – that manifests this all-important systematicity. It is pleasing to begin with such an architecture: it is neat, all-encompassing, and one can get down to the task of figuring out how persons learn it and exercise their knowledge of it. But there are, as we shall see (and as many have already pointed out) great costs with this way of proceeding to understand how the mind learns and works.

A great deal depends on where one begins. Imagine that you are interested in explaining how persons learn to speak and do in fact exercise their ability to speak a language. One way of beginning here is to first look at the body of language itself. As you do so, you might very well feel overwhelmed by the variety of uses and expressions, so you look for a way to make this more manageable. As you do so you discover, happily, the rules of grammar. This makes your work much more efficient: what you do now is define the knowledge of English as a matter of the knowledge of such rules (where you encounter difficulties with the scope or diversity of the rules, you posit more general, underlying principles from which they flow). This way, whenever you encounter a specific example of learning a language and speaking it, you explain it by looking at the expression itself and analysing it as an instance of an

application of some rule (again, if you cannot find a rule in the textbooks or in the linguistics literature, you can argue it is implicit, or that there is some higher order principle that dictates it). You still need to show how the rules are learnt and applied in practice, and given that you do not want to make it the case that every time we speak a language we turn to a rule book or try to remember the relevant rule, you look for some other mechanism (somewhere in the brain, perhaps) that can run the program (apply the rule) for us. Certainly, you will then need to confront the question of how the rules are there already in the brain, and to do this you may posit some evolutionary or innate mechanism that explains their presence (not only in that one individual, but in everyone: a kind of universal innate or evolutionary grammar embossed on the mind, and one that explains the terrific variety and infinite of possible expressions).[116]

But now let us try to begin elsewhere. Imagine you are an infant and your parents are talking by the bedside. They are talking about what they did that day: 'how was your day, sweet sugar pie'? etc. As you listen, you get exposed to hundreds of sounds – later, you will be exposed to thousands, hundreds of thousands, and eventually millions. It is possible that as you listen you begin to correlate sounds – such-and-such a sound associated with, say, speaking about the past. Your mind – being an infant you have a supple, flexible mind that can soak up sounds at an incredible rate – actively and creatively makes links between sounds and situations / circumstances / tasks, some of which it may lose along the way (as not effective) and others that it may keep. Indeed, it keeps doing this, over and over again, making links and connections; marvelling at sounds it has never heard, playing with them – it is active, learning, creating, though sometimes precisely by listening, soaking in, being exposed to sounds. At some point, some years later, you may enter a school and as you do so persons around you (friends, teachers, parents) may correct you (sometimes on the basis of rules, and sometimes because something does not sound quite right to them), and your learning may become more disciplined, more systematic, such that you learn rules and styles, genres and forms. These devices for systematising and

---

[116] Here is how Chomsky puts it: 'To begin with, let us assume that it makes sense to say, as we normally do, that each person knows his or her language – that you and I know English, for example – that this knowledge is in part shared among us and represented somehow in our minds, ultimately in our brains, in structures that we can hope to characterise abstractly, and in principle quite concretely, in terms of physical mechanisms' (Chomsky 1980, 1-2). As was noted above when we discussed Chomsky in more detail (see Section 8.C.), the concept of learning gets lost on this way of seeing linguistic ability: we do not learn to speak a language, so much as the rules 'grow' in our minds.

justifying your ability to speak will help you – but not help you to the extent that they overtake your previously acquired experience (all the links and connections you made) and not, then, such as to dominate your ability to play with those links and connections (invent new uses, new ways of expressing yourself, your own 'voice').

These are two very rough descriptions, and they are not offered as theories, but what they do reveal are how two different beginnings can produce two very different attitudes to rules: in the one, rules are what define the very object of explanation (knowledge of language), thereby becoming the very thing that must be always prioritised as the driving force in learning and speaking a language; in the second, rules are one way (often a very important and efficient way) for a community to discipline itself (for certain purposes, e.g. educational, professional), and / or for a person to sometimes guide themselves or to criticise herself or others. On the second view, then, rules become a resource capable of being an efficient means for certain ends, but by no means the explanatory target or the basis of explanation of any instance of learning and speaking.

Here you might ask: but what is the advantage of the second approach? You might add that it seems to explain less, or that it 'gives up' just at the moment when we might need an extra effort to show how 'the rule' really is there and was there all along, and how it does cover the new use or all these seemingly different looking uses. The reply to this is that the advantage consists in how it pictures the mind learning and working: what it leaves room for, that the first does not, is how actively the mind can itself create links and connections, including how it can do so uniquely and distinctively, while nevertheless potentially and contingently overlapping with how others create links and connections. It is true that this does not allow for a neat, all-encompassing, somehow already complete (at least in the abstract form of a system of rules) target and basis of explanation; but it does allow for systematicity to be attempted and appealed to when this matters (e.g. when devising a curriculum) – a kind of 'pop-up' or 'ad hoc' systematicity – and it combines this with respect for variety and change in learning and speaking a language. It may be less satisfying in some general theoretical way, but then if it is, this need not be a disadvantage of the picture – instead, it may indicate something (perhaps something we would prefer not to learn) about the practice of theory (or one kind of practice of theory).

The above are very general remarks. Let us see how they might manifest themselves in more detailed consideration of proposals made by theorists. We begin

first with the Chomsky-inspired Pinker take on linguistic ability, and then move on to an argument made in cognitive science on behalf of the role of rules in reasoning. We then finish with some remarks about 'the frame problem' in cognitive science.

A. *Pinker's System of Rules and Exceptions*

We referred above to Chomsky and mentioned that one of the influential contemporary advocates for his view (though one that argues for an evolutionary rather than simply innate origin of universal grammar) is Steven Pinker. In an early paper, 'Rules of Language' (1991), Pinker helpfully outlines certain key features of his approach. The paper is instructive, also, because it tries to articulate a kind of rapprochement between the computational rules-and-representations picture of learning and speaking a language (which Pinker is personally most attracted to) and what he calls an 'associationist' picture, where the emphasis is on a 'network of interconnected units modified by a learning mechanism that records correlations among frequently co-occurring input patterns' (Pinker 1991, 530). Pinker's strategy is to argue that although the associationist picture can help explain some 'components of the language system', it does not explain many others, and these others it does not explain are precisely the ones that are explained by the computational picture.

Pinker's argument is based primarily on the idea that whereas the computational model can explain our acting in accordance with the rule for transforming regular verbs into the past tense, it does not deal with those irregular ones (which simply need to be learnt by 'associative memory'). This is how he puts it: '[R]egular past tense forms are computed by a rule that concatenates an affix with a variable standing for the stem. Irregulars are memorised pairs of words' (Pinker 1991, 531). It is difficult to see how the 'memory' referred to in the context of irregular verbs is 'associative': it reads, instead, like a case of simply memorising that which does not conform to the rule. Pinker does allow for that which is memorised (the irregular form) to be stored (obviously, for otherwise it would not be memory) such that it could generate an extension (e.g. we could learn that *string* and *strung* go together, and this may mean we will transform *bring* into *brung*), but this is hardly an associative extension – indeed, it is nothing but the application of a rule (e.g. when a verb ends with –ing, then use –ung for the past tense). In short, Pinker's 'associative memory' is not associative at all, but simply a very limited rule (perhaps Pinker does not think of it as a rule because he thinks of it as only applying to one case). Notice,

further, that the 'associative memory' (which is not really associative) is depicted only as producing negative results (*brung* is not how we usually transform *bring* into the past tense), whereas that which produces positive results (basically that which can be generalised) is said to be the production of the rule (and the computational model). On this picture, then, the associative model is completely lost sight of, for it is restricted to the mere memorisation of irregularities (which cannot be generalised – or when they are attempted to be generalised indeed only reveal their own limitations). The point here is that Pinker's attempted rapprochement (though it is anything but, for it loses sight of the associative picture) between the computational and the associative does not work because he already sets up the target of explanation by defining it on the basis of a rule and exceptions (the latter being single uses that do not conform to the rule). The associative story is not allowed to get going (for instance, the child is not depicted as possibly learning to associate most past tense verb forms with the suffix –ed) – instead, it is banished to memorising exceptions (that the computational side must even discipline, i.e. not allow it to generalise).

Has Pinker's picture changed in later work? Consider a 2001 paper entitled 'Words and Rules' (a short summary of Pinker 1999). Here, Pinker presents associative memory as working in the following way: 'The word duck doesn't look like a duck, walk like a duck, or quack like a duck, but I can use it to cause you to think the thought of a duck because all of us at some point in our lives have memorised an association between the sound and that meaning' (Pinker 2001, 14). So now, rather than explaining how we learn exceptions to rules, associative memory is said to do the work of linking certain sounds to certain meanings. What about rules, and the computational model? Here, Pinker gets into his stride: 'But of course we don't just blurt out individual words' – as we would, on Pinker's view, if we only had associative memory – 'We combine them into phrases and sentences, and that brings up…combinatorial grammar – what Wilhelm von Humboldt called "the infinite use of finite media"' (Pinker 2001, 14). It is the computational model – or rules, in short – that allow for the production of infinite combinations (i.e. infinite applications of the rules), e.g. '...because our knowledge of language is couched in abstract symbols – noun, verb, subject, object – the same rules allow us to talk about a big dog biting a man and a big bang creating a universe' (Pinker 2001, 14). Here, it is the computational side that is said to rescue our creativity, though our creativity is restricted to making infinite numbers of combinations that conform to rules.

Much of this view depends on a certain understanding of rules, i.e. that they are formulated in a finite way, but nevertheless apply infinitely. The problem with this view is precisely the second step: infinite application (the very existence of 'exceptions' helps us see the problem too). Appeals to infinite application assume that the rule, as we know it (in its finite formulation), is automatically relevant to the new case. Return for a moment to the alleged rule for regular verbs transformed into the past tense (e.g. walk-walked, jog-jogged, kiss-kissed: Pinker 2001, 15 – the same examples are drawn on in the later paper as in the earlier 1991 paper). Consider being faced with a new word (the following is Pinker's own example): 'In 1958, Jean Berko Gleason brought some four year-olds into the lab and said, "Here is a man who knows how to wug. He did the same thing yesterday. He…….."' (Pinker 2001, 15). Pinker uses this example to illustrate the 'productivity' of the rule (for regular verbs, i.e. add –ed), so that we get 'wugged'. But what Pinker does not say is why 'wugged' is to be assumed as the use that we, in the community, will adopt or should adopt. In other words, the 'infinite application' of the rule is only a presumption.[117] The community, faced with 'wug', might also have decided on a different suffix, e.g. 'wugook', as in 'forsook.' Of course, if we did adopt 'wugook', Pinker might reply: but this is just an exception – it does not threaten the 'infinite application' of the rule. The reply does not work, for it undermines itself: the very presence of 'exceptions' shows the limits of the rule (these being limits that are not set in advance). Of course, we can call it 'an exception' to try to save the appearance of the infinite application of the rule, but all we are doing, in effect, is saying that some pattern that we rely on a good deal may usually be a good guide as to what to do with a new word (it is such an important resource for us that we tend to elevate it to the status of a rule with infinite application). Notice that it is also not a reply here to say that the rule is finitely applicable to that which it is applicable, i.e. that it generates those cases that fall within its scope, for this is just to repeat the limitation of the rule. Indeed, the rule can be contingently used as a guide (as with any normative resource) because it is limited;[118] to speak of infinite applicability, or of a rule generating infinite uses, is to blow smoke in front of our eyes – to make us forget about the instance of being faced

---

[117] The way Pinker is using rule here suggests a prediction which, if satisfied, would hardly warrant the conclusion that a rule was being followed. This is because for rule-following one needs more than rule-compliance (one needs evidence of some kind of mediating and / or critical reflective attitude).

[118] Another way to put this point is to say that the applicability of the rule is limited, but in an indeterminate way (and this is because we need to decide when it is relevant).

with a new word and thus the instance where we are confronted with the limits of (or, sometimes, depending on how we relate to the new word, the irrelevance of) the rule. We can be very impressed with a 'rule', for we may use it to simplify our reply to new words (e.g. we could just tell ourselves to treat anything new as either a violation of – and thus, on Pinker's view, an exception to – or an instantiation of the rule),[119] but what is forgotten about here is that the very applicability of the rule to a new word depends on an active and creative evaluation as to the relevance of the rule, it being possible that the rule will be irrelevant and that we (potentially) come to articulate a different set of rules ('potentially' because we could leave the matter unresolved). We miss the importance of this active and creative exercise of relevance – and thus of a non-rule guided, indeed non-guided form of cognition – when we begin with, and subsume everything, under some system of rules and exceptions.

B. Cognising Rules

Let us move a little away from the Chomsky / Pinker approach to language now, and consider some other defences of computation (or the allegedly necessary role of rules in cognition). A useful paper that we can consider here is 'The Case for Rules in Reasoning' by Edward Smith, Christopher Langston and Richard Nisbett (1992). The paper is useful because it sets out eight criteria 'for establishing whether reasoning makes use of abstract rules' (Smith et al 1992, 7). Here are the eight criteria in summary:

> Criteria Stemming from Linguistics
> 1. Performance on rule-governed items is as accurate with unfamiliar as with familiar material;
> 2. Performance on rule-governed items is as accurate with abstract as with concrete material;
> 3. Early in acquisition, a rule may be applied to an exception (the rule is overextended);
>
> Performance Criteria
> 4. Performance on a rule-governed item or problem deteriorates as a function of the number of rules that are required for solving the problem;

---

[119] Certainly, we can tell ourselves this, and we can try to behave as if we, by necessity, are guided in this way, but try as we might, the activity and creativity of our mind will outstrip such a commitment. No commitment, or any conservative normative attitude, can control the mind's ability to create relevance.

5. Performance on a rule-governed item is facilitated when preceded by another item based on the same rule (application of a rule primes its subsequent use);
6. A rule, or components of it, may be mentioned in a verbal protocol;

Training Criteria
7. Performance on a specific rule-governed problem is improved by training on abstract versions of the rule;
8. Performance on problems in a particular domain is improved as much by training on problems outside the domain as on problems within it, as long as the problems are based on the same rule.
(Smith et al 1992, 7)

Notice that these are not so much criteria for determining when reasoning makes use of rules, but more a matter of defending the case for rules in reasoning, i.e. defeating arguments that try to show that rule-following reasoning is not (e.g. when faced with unfamiliar, concrete material) at work. As the authors say, 'each criterion essentially embodies a phenomenon that is more readily explained by a rule-based approached than by an alternative model' (Smith et al 1992, 2). Let us, then, examine, briefly, each criterion.

The first criterion is that 'performance on rule-governed items is as accurate with unfamiliar as with familiar material'. Everything here turns on what is understood by the familiar and the unfamiliar. By familiar, Smith et al appear to mean familiar content. Thus, they give the example of the *modus ponens* rule, If p then q; p; therefore q, and their unfamiliar example is: 'If gork then flum; gork; ?', suggesting that 'when presented with this time, you might code it, in part, as an "If X, then Y" item type. This would suffice to access *modus ponens*. Next, you would instantiate p with "gork" and q with "flum." Then you would apply the rule and drive "flum" as an answer' (Smith et al 1992, 8). This is a rather remarkable example because it presents what is billed as the unfamiliar in the form anticipated by the rule (in the 'If…then' form). Surely, if we are to use the unfamiliar as a criterion for when rules are relied on – or to what extent they may be relied on – when reasoning, we will need to give examples of the unfamiliar that do not take on the familiar form of the rule (and not just some unfamiliar content). In other words, the rule is identified / characterised by the form, and not the content – so what one needs by way of examples of the unfamiliar is material that is not yet formed in accordance with the requirements of the rule. In the absence of examples from Smith et al about the unfamiliar in this sense, we must conclude that the first criterion is not a success.

The second criterion is that 'performance on rule-governed items is as accurate with abstract as with concrete material.' As with the previous criterion, a lot depends on the definition of 'abstract' and 'concrete'. It seems that, once again, Smith et al define these in ways that presuppose the form of the rule. Thus, 'abstract' material is, for instance 'If A then B; A; therefore B' (Smith et al 1992, 15; 'abstract material' here is basically a repetition of the rule itself, but with different letters: A instead p etc), whereas 'concrete material' is for example, 'If the letter is L, then the number is 5; the letter is L; Therefore the number is 5' (Smith et al 1992, 15). The way that Smith et al rely on this difference is that they say that when the latter was tested (it was tested by having participants evaluate the conclusion as either valid or invalid), all the participants said it was valid – and thus, according to Smith et al, this showed that concrete material does not make a difference as to whether a rule is used or not. Again, this conclusion does not follow: first, because there is no difference between abstract and concrete material, for the abstract here just is the rule, and the concrete but an application of the rule; and second, because 'concrete' is already presented in the rule form.

Criterion 3 states that 'Early in acquisition, a rule may be applied to an exception (the rule is overextended)'. This criterion wears its problem on its sleeve: it assumes that what was learnt is a rule, and it assumes that we know, already, when a rule is properly applied and when it is not (i.e. what is an exception and what is not). Smith et al's example is the one we have met above in discussing Pinker: 'studies of how children master the regular past-tense form of English verbs' (Smith et al 1992, 16). Smith et al say that studies have shown that children often use the same tense form, i.e. adding –ed, 'even though the items should have been marked as exceptions' (Smith et al 1992, 17). According to Smith et al, this shows that a rule is being followed. But it is hard to see how this conclusion follows. On this view, any habit could be a rule: imagine that I have always heard anyone speaking of the past to use the –ed form for verbs. It is hardly surprising, in such a case, that I am likely, every time I face the task of transforming a verb into past tense, to use the –ed form for verbs. But does this mean that I must have learnt a rule to that effect such that I am applying it in this instance, or there might be another story as to how I developed an ability to tell a story in the past tense? Smith et al might reply here that 'over-extension' shows they are following a rule, but this depends on how one understands 'over-extension'. 'Over-extension' only counts as so from the perspective of

someone, e.g. a grammarian, who has come across the particular verb in question, and who has classified it as an exception (and here, as has been argued above, we need to examine the process of how something might be classified as an exception). In the absence of a grammarian amongst the children, it is difficult to see any of the children standing and saying: 'no, that verb is an exception; do not apply the rule'. If they did say this, there might be a case for saying that, on this occasion, that child had learnt and was following the rule for transforming verbs into –ed form (but notice that it assumes that the child is already familiar with the classification of that verb as an exception). But Smith et al give no evidence of such criticism articulated on that basis: all they say is that if the children 'over-extend', it shows that they are relying on a rule – and, as has been argued here, this is just to say that Smith et al assume that the children are following a rule and that Smith et al give us no reason to accept this assumption.

The fourth criterion shifts gear and introduces the category of 'performance criteria.' This fourth criterion is: 'Performance on a rule-governed item or problem deteriorates as a function of the number of rules that are required for solving the problem.' At first glance, this criterion promises to reveal that there are examples of problems that are so complex – would require the application of so many rules – that it is unlikely they are solved on the basis of rules. But this is not what Smith et al mean at all. What they mean, quite simply, is that it might take longer for us to compute the rules where more than one instantiation of the rule needs to be computed. In other words, Smith et al assume that the problem in question is a problem that can only be solved by following the rules; the only issue is whether us having to apply (say) the same rule twice means that our performance (how fast we solve the problem) may suffer. At one point, Smith et al say that some studies show that in some cases of problem solving, i.e. cases where rules are frequently applied in succession, these rules may be 'compiled' or 'chunked' into a single rule (Smith et al 1992, 19) – this is especially after some passage of time (at first I solve it on the basis of those rules, but then I 'chunk' those rules into one single rule). Notice how even this case is described: the shortcut used (the ability developed to solve the problem) is still one that must be characterised in the form of a rule (just a single rule, rather than a succession of rules). No awareness is shown here for the possibility that when I learn to solve a problem that is as complex as that, my ability to solve it may have nothing to do with either computing those rules very quickly or computing some

single 'chunked' rule. The trick here is how one characterises the problem: if one thinks of 'the problem' as a matter of reaching this and only this result by following these and only these means (i.e. these rules), then one will be forced into finding a description that says we either learn to compute those rules very quickly or that we somehow chunk those rules into one big rule. One will only define 'the problem' in this way if one presupposes that the only way of solving it is by following the rules. It is surely very unlikely that all problems human beings solve are ones where the result is tightly connected to the means that there is no other way to solve it (especially when one gets 'good' at solving it – here, getting 'good' at solving it need not mean getting 'good' at computing the rules quicker; indeed, relying on some fixed set of rules to solve problems may be a sign of a novice: the point is that being 'good' at something, having an ability, is not reducible to being a more efficient computer – it only seems that way if one defines the domain as one of these and only these problems, and these and only these means for solving them).

The fifth criterion is as follows: 'Performance on a rule-governed item is facilitated when preceded by another item based on the same rule (application of a rule primes its subsequent use).' This criterion is not helpful in determining whether a piece of reasoning is following a rule or not; thus, even though it may be the case that persons who have just been told (or flashed) the *modus ponens* rule are more likely to think that that rule will be relevant to solving a problem they are then given, by no means is it the case that they necessarily will solve that problem on the basis of that rule. Notice that not only does the priming method not guarantee the person will rely on the rule in solving the problem given to him or her, but that it may also blind researchers into thinking that there must be some way (if the person solved the problem, and the problem could be solved by relying on the rule) that the person relied on the rule (rather than considering an alternative means). Of course, if the researchers define 'the problem' as one that can only be solved on the basis of the rule, then, together with priming, they simply reinforce their bias for finding rules where none may be there to be found.

The sixth, and final 'performance' criterion, states that 'A verbal protocol may mention a rule or its components.' We have already – in looking at Stueber's proposal – given examples where there need not be any necessary connection between how a person does something and what he says about why he did or even how he criticises himself. Whereas many theorists have recognised that such connections ought not to

be assumed, Smith et al say that 'Presumably, the protocol is a direct reflection of what is active in the subject's short-term or working memory…and if a particular rule has been in working memory, then it may have been recently used' (Smith et al 1992, 20-21). Certainly, it is possible that someone who reports having relied on a rule did in fact rely on it to solve some problem – but it is also possible that they did not; there is no reason here for a presumption about direct reflection (in fact, as was discussed in Section 10.D., there is reason against such a presumption). Smith et al's bias to rules is also evident in how they articulate a limitation of this criterion: 'the protocol criterion is of limited diagnosticity, given that there are cases of apparent rule following in which the rules cannot be reported (namely, in language)' (Smith et al 1992, 21) – again, this simply assumes that we follow rules when we learn / speak a language despite not being able to 'report' them.

The last two criteria are 'training' criteria. Both of them need not be discussed, for they are problematic in their formulation. The seventh criterion, is as follows: 'Performance on a specific rule- governed problem is improved by training on abstract versions of the rule.' The last and eight criterion reads: 'Performance on problems in a particular domain is improved as much by training on problems outside the domain as on problems within it, as long as the problems are based on the same rule.' Both criteria assume that – rather than test whether – 'the problem' is a 'rule-governed' one, or one 'based on the same rule'.

We have worked through each of the criteria and have shown that Smith et al make not a case for rules in reasoning, but a case for seeing how much they are determined to see (or, better, presume the existence of) rules wherever they can.

C. The Frame Problem

Part of the difficulty with debates over the scope and viability of the computation model is that almost anything that is contrasted with rules can be forced into a rule-based straightjacket, i.e. can be described in terms of rules (which makes it look as if it was only a rule-based computation after all). As Ulrike Hahn and Nick Chater (1998) put it, the form of a rule 'can be used to include procedures for computing similarity' (Hahn and Chater 1998, 200), or for 'banks of connectionist units' or for 'rules of thumb' (Hahn and Chater 1998, 201), and so on, and this is especially so when one begins to speak of 'implicit rules.' Hahn and Chater tackle this tendency by saying that 'rule- and similarity-based processes differ regarding the way the

representation of the new item is integrated with existing knowledge', or more succinctly, differ in terms of 'the way in which representations are used' (Hahn and Chater 1998, 201). This strategy allows Hahn and Chater to reverse some common assumptions, e.g. that 'evidence for highly abstract mental representations' is evidence for rule-based processing – as they point out, by no means is it the case that similarity-based processes cannot be utilised for abstract representations; rather, they offer one way of understanding how we cope with abstract representations (e.g. by partial matching of past with present, rather than strict matching as is the case, according to Hahn and Chater, with rule-based processes: Hahn and Chater 1998, 202). This reversal is a helpful one, but it is not clear whether it ought to lead us to think that rule-based processes are necessarily a matter of strict matching (as Hahn and Chater argue it does); advocates of rules might say that rules themselves may be used in different ways, or that 'rules' is too general a term and where one needs something like partial matching one might speak of 'principles' or 'standards' or normative resources that allow for more relaxed matching. It is difficult to orient oneself in this debate because the terms are being used so loosely, and almost everyone in the debate is using them differently, or re-defining them.

But the debate does help us to get at several points we can make about the reference to rules in this context. There is, it seems, something about the general form of a rule ('in such circumstances, do this or that' – or, 'if you find yourself in these circumstances, then you ought to do this or that') that lends itself to being used as both a description and a prescription. The form of a rule is somehow both backward and forward looking, i.e. it promises a description of what we have learnt, but also some sort of recommendation or warning about what to do in the future (whenever you encounter such-and-such, do or do not do such-and-such). It may be this two-directionality of the form of a rule that makes it so flexible, i.e. that it can be used as a description and as a prescription. But this two-directionality also has a risk in that we will tend to think that a rule must always both be a description and a prescription, i.e. that it both prescribes what ought to be done and that (because we perceive persons as acting, say, in some way that we expect them to act) it is also a description of their behaviour. There is a sense here in which what we use as a guide for ourselves is projected onto understanding the behaviour of others, but with the crucial difference that we forget we are projecting (just because the rules matter so much to us). Perhaps this same point can be put differently, namely: the tendency to describe what we have

learnt in the form of rules is accompanied by a tendency to conserve what we have learnt (as we have described it), e.g. by presenting rules as finite enough to guide us, but infinite in application, or by forgetting the first instance in which something needs to be classified as either within the scope of the rule or not (e.g. an exception or a new rule), or by positing implicit rules (or rules that are somehow there and always were even though we did not know they were), or by assuming the rules apply themselves or are automatically relevant, and so on. As a result, what are undoubtedly important resources for the mind are used as description of how the mind (always and necessarily, by defeasible regular default) learns and works. The cases are all described in such a way as to make it seem as if there was no other option but to follow the rule, or where not following the rule is obviously a mistake (precisely because it is a violation of the rule). The final outcome is one where it appears difficult to see how the mind could learn something anew for the first time, or do something that is not either in accord or in violation of some rule it already has (perhaps without knowing it). One must wonder, when seeing this, whether it is not better to resist the tendency to want to describe what we have learnt, or how the mind learns and works generally, in the form of rules, but to look for an alternative description.

Notice that the limitation of a rule-based model of cognition – computation, as we have been looking at it here – is not solved by acknowledging that there are exceptions to rules and that the recognition of these exceptions may not be a matter of rule-following (this is the line taken by Pinker). As we have shown above, this is because we are not faced with a *fait accompli* as to whether something is within the rule or an exception; instead, we actively and creatively engage with what we encounter, potentially deciding some rule is applicable or not applicable, or indeed creating something as relevant and important that was not dreamt of by the rules.

This issue of relevance is an important one. The most prominent way in which it appears in cognitive science is as 'the frame problem',[120] which is typically understood as the issue of how it is that the mind decides (for many, the challenge is to describe how the mind decides without us being aware) that some rule is relevant. There have, in other words, been many attempts at solving the frame problem – but what 'solving' it is often said to mean is that we need an account that somehow fixes

---

[120] For a discussion of this problem in the context of moral normativity, see Horgan and Timmons 2009.

the relevance of the rule, e.g. Jerry Fodor's (1983) notion of modules may be understood as precisely an attempt to fix relevance (in such-and-such a domain, such rules are relevant; hence the description of the mind as modular, i.e. as somehow built to enable the determination of relevance).

The point that needs to be made here is that this is not to 'solve' the frame problem, but just to push the problem further back (e.g. with respect to Fodor's modularity thesis, we can ask: but why is it that this domain – accompanied by these rules – is relevant?). The point is that the frame problem is not a problem that can be solved: the rules, as we have observed many times in this thesis, do not apply themselves, or better, do not make themselves applicable automatically (they are never automatically relevant). The frame problem is better understood as a marker of the limitations of a rule-based account of cognition (and, simultaneously, also the limitations of an understanding of relevance as something that can be determined). One can go a long way with a rule-based description of cognition if one assumes that, for certain tasks / domains / activities etc, certain rules are unchangeably relevant. It might seem like in many cases such an assumption will be a safe one (e.g. when playing chess, it seems safe to assume we ought to learn and follow such-and-such rules of chess). But the problem with it is that it all too easily results in a description of the mind passively learning what is already there, out in front of it, ready to be learnt, and doomed to re-enacting what it learns, and it forecloses the opportunity for describing how the mind may learn things differently, including more tentatively (not describable as a matter of learning to apply, by default, some general rule). In essence, the frame problem shows us that we can all too easily become over-impressed with (and forget about the limits of) resources (in the form, say, of rules) that we ourselves may contingently use to guide ourselves and / or to criticise ourselves or others. When we forget, in this way, the contingent usefulness of normative resources, we also make it difficult, if not impossible, to see and describe the full array of abilities we need to participate in social life, as that participation is understood from the second-person perspective. Our cognition – including its normative aspects, i.e. our ability to direct ourselves to do something we feel we ought to, or our ability to evaluate something as possibly appropriate or inappropriate – is not dependent on the learning, internalisation and thereafter default following of rules.

Chapter 12. Necessarily Conceptual

The most prominent advocate of the thesis of necessarily conceptual cognition, and also one who argues that this conceptualism is explicable on the basis of rules and thus rule-following (and that this is also what captures its normative dimension, without which it would not be uniquely human cognition), is John McDowell. There are few contemporary bodies of work more influential, but also more difficult, than McDowell's. The structure of our discussion in this Chapter will be as follows: first, to consider *Mind and World* (1996), and to evaluate what kind of picture of the mind learning and working we get; and second, to consider aspects of the recent debate between McDowell and Dreyfus, which will allow us not only to further our inquiry into McDowell's view, but also to introduce themes that emerge from Dreyfus's work (which will offer a useful bridge to material discussed in Chapter 13).

A. Mind and World

That one of the key problematics in *Mind and World* is the 'demarcation problem' – i.e. how human beings are different to 'brutes' or 'mere animals' – is clear right from the beginning of McDowell's text. McDowell pursues the distinctiveness of human beings by referring to them as 'rational animals', and one of the main purposes of the book is to show in what sense human beings are essentially or necessarily rational.[121] The precursor here, in more ways than one, is of course Kant. But perhaps the key figure – the key mediator between Kant and modern analytical philosophy (or at least the line represented today by McDowell and Brandom) is Sellars, whose image of 'the space of reasons' is a leading image in McDowell's text. However, as important as Sellars is for McDowell, in the 'Introduction' to the second edition (1996) of *Mind and World* McDowell asserts that his own position differs from Sellars's in an important respect. McDowell quotes Sellars as saying (in 'Empiricism and Philosophy of Mind': Sellars 1997 [1956]): 'In characterising an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says' (Sellars, quoted in McDowell 1996, xiv). McDowell then distinguishes his own project from this statement by saying: 'It is a way of repeating what I have just

---

[121] In his as yet unpublished manuscript, 'Philosophy of Action', McDowell acknowledges that the theme in *Mind and World* was the discontinuity between human beings and animals, and in the manuscript he tries to emphasise more of the continuities.

been urging…to say this: though Sellars here speaks of knowledge in particular, that is just to stress one application of the thought that a normative context is necessary for the idea of being in touch with the world at all, whether knowledgeably or not' (McDowell 1996, xiv). The reason why this is so representative of McDowell's project – in a way that matters greatly for this thesis – is that McDowell effectively extends the context of justification into all possible tasks the mind might undertake. In doing so, he inherits all the problems we have been discussing, especially: how could I be said to act in such a way that I am always acting on certain grounds that I can then cite (if required to cite) as the justification of my action? The setting up of this problem, and the pursuit of an answer, is one of the foundational moves of the governance view – for, from there, it is only a small step to the idea that the content of the reasons in question is given by rules (for instance, via concepts defined as rules).

The prioritisation of the context of justification goes hand in hand with another matter that is emphasised in *Mind and World*: the idea that the mind is capable of being directed at how things are. When McDowell asks: 'How is it possible for there to be thinking directed at how things are?' (McDowell 1996, xii), he is asking, effectively (given his project / how he sets things up): how is it possible to think at all; how is thought possible? This can be put another way: 'thinking', says McDowell, 'aims at judgement, or at the fixation of belief' (McDowell 1996, xiii). If thinking was not a matter of aiming for a justifiable judgement as to how things are, then it would not be thinking. Indeed, the very notion of 'rational animals' communicates this as well: we are only different to animals because we are capable of being rational, i.e. we are capable of thinking in such a way that aims at justifiable judgement as to how things are. As we shall see in a moment, according to McDowell, aiming at justifiable judgement as to how things are is itself expressible only in the language of concepts: of our faculties being essentially or necessarily conceptual – in other words, aiming at justifiable judgement as to how things are just is to be guided by concepts. Our 'rationality' (in that above sense) is our 'second nature': 'Human beings acquire a second nature in part by being initiated into conceptual capacities, whose interrelations belong in the logical space of reasons' (McDowell 1996, xx). As we shall see in a moment, tying thinking to concepts, and defining concepts in terms of rules, also helps to set up the possibility of a standard by which persons' judgements can be said to be either correct or incorrect (and at first blush objectively so – for they would be judged on the basis of the rules; i.e. it is only

by virtue of the applicability and the application of some rules that persons could be said to be correct or incorrect in the desired manner, the manner dictated by what objectivity demands).

What are concepts and why is it that human cognition is necessarily conceptual? 'Concepts', says McDowell, 'mediate the relation between minds and the world' (McDowell 1996, 3), and one way to argue for them being unnecessary is to argue that we have no need for mediators, although we can distance ourselves from how things appear to us by utilising them (as when we might, for instance, try to 'stand back' and 'figure out' what we are experiencing). But McDowell does not confine his thesis about human cognition to an experience of distancing – he asserts, much more strongly, that our relation with the world is necessarily mediated by concepts, whether when distancing or not. This is so, once again, because it is, according to McDowell, only by virtue of being guided by concepts that we can make 'warranted judgements about the world' (McDowell 1996, 5), and if we could not make warranted judgements about the world (here, 'about the world' should presumably be read as 'about how things are') we would be merely animals, and not rational animals. The possibility of warrant, then, is in a way definitional of concepts – it is what characterises them as the things they are, i.e. that they can provide, or that they are the means to reaching, warranted judgements. McDowell puts this in constitutive language: 'we took this possibility of warrant to be constitutive of the concept's being what it is, and hence constitutive of its contribution to any thinkable content it figures in, whether that of knowledgeable, or less substantially justifiable, judgement or any other' (McDowell 1996, 6).

There is another way of putting this, which refers to needing / wanting reassurance that things are the way they appear to us, where it is thought that that reassurance can only be provided by something external to us (some rules) that tell us when our perception of things (our experience of things appearing to us in a certain way) are correct or incorrect: 'What we wanted was a reassurance that when we use our concepts in judgement, our freedom – our spontaneity in the exercise of our understanding – is constrained from outside thought, and constrained in a way that we can appeal to in displaying the judgements as justified' (McDowell 1996, 8). Insofar as this is the approach, we can ask: why think that whenever we experience the world, this experience must be marked by the need to reassure ourselves that things really are as they appear? One feels the hot breath of the sceptic down the neck of *Mind and*

World at every turn: this desire to respond to the sceptic is there always in the background, encouraging McDowell to picture the mind as always and necessarily acting on some ground (such that it can be unchallengeably correct or incorrect in doing so). The question that emerges is: must the sceptic be silenced, or can the sceptic be seen to be making an important contribution, but not so important as to make us depict the mind as perennially in need of reassurance that what it is experiencing is how things really are?

But let us go more slowly: what exactly does McDowell say about how concepts mediate our relation to the world? Here, McDowell's answer, roughly, is this: 'empirical substance is transmitted from the ground level to empirical concepts that are further removed from immediate experience, with the transmission running along channels constituted by the inferential linkages that hold a system of concepts together' (McDowell 1996, 7). There are several things here that demand further explanation: e.g., what are 'empirical substances'? In what sense are concepts 'empirical' (is there another kind of concept that is not empirical)? Why are concepts not enough – why does one also need there to be a system of concepts linked by inferential linkages? What are these 'inferential linkages' exactly?

Part of McDowell's strategy in clarifying his own position (and answering the above questions) is to contrast his position with that of the 'Myth of the Given' (a phrase used first by Sellars). The reason this is so is because the Given is precisely that which does not give us the reassurance we (according to McDowell) are after:

> The Given is the idea that the space of reasons, the space of justifications or warrants, extends more widely than the conceptual sphere… But when we make out that the space of reasons is more extensive than the conceptual sphere, so that it can incorporate extra-conceptual impingements from the world, the result is a picture in which constraint from outside is exerted at the boundary of the expanded space of reasons, in what we are committed to depicting as a brute impact from the exterior. Now perhaps this picture secures that we cannot be blamed for the inward influence of what happens there. What happens there is the result of an alien force, the causal impact of the world, operating outside the control of our spontaneity. But it is one thing to be exempt from blame, on the ground that the position we find ourselves in can be traced ultimately to brute force; it is quite another thing to have a justification. In effect, the idea of the Given offers exculpations where we wanted justifications. (McDowell 1996, 7-8)

This is an important passage that we need to pause over. McDowell certainly has a point here about what can serve as the kind of standard that can yield a justification (as opposed to a mere exculpation). But the difficulty is that we can speak of experience in two different ways here: we can speak of it from the perspective of wanting reassurance or justification, or we can speak of it from the perspective of not needing to worry or feel anxious about justifying / being able to justify ourselves. Experience becomes the 'myth of the Given' when it is asserted (and McDowell may be right to criticise this assertion) that there is some sort of basic view of experience that can serve the interests of justification, but he misleads us if we conclude from this (as he seems to conclude) that no experience is possible outside this context of justification; in short, the context of justification does not have the jurisdiction – the authority, if you like – to dictate our understanding of experience. The matter can be put another way: experience without the Given, and thus not evaluated from the perspective of an interest in justification, is something that does not care about (hence feels no need for reassurance) what the philosopher trying to answer the sceptic cares about, i.e. being able to know whether our judgements about how the world are really do correspond to how the world is (or, put slightly differently, of us having access to a way of articulating the way the world is such that we can evaluate whether someone thinking about how the world appears to them is either correct or not correct). So read, the context of justification makes us neglect the possibility of experiencing in ways that are not marked by anxieties / doubts injected by the sceptic (put differently, in seeking to answer, once and for all, the sceptic, we squeeze all of experience into a mould that can provide justification). Again, the point of relevance to this thesis is that when we do prioritise the context of justification, one of the things we can do is miss (fail to pay attention to, neglect the importance of) a way of experiencing (e.g. attending silently to, and appreciating) that cannot be subsumed under its contribution to (or lack of) justification or warrant.

Placing experience in the context of justification also has another effect that reappears throughout *Mind and World*, i.e. the idea that 'experience is passive' (McDowell 1996, 10). McDowell has quite a specific notion of passivity. He clarifies, for instance, that it does not mean 'to deny that experiencing the world involves activity. Searching is an activity; so are observing, watching, and so forth' (McDowell 1996, 10, fn. 8). What he means, instead, is that 'In experience one is saddled with content' (McDowell 1996, 10), i.e. 'one's control over what happens in experience

has limits: one can decide where to place oneself, at what pitch to tune one's attention, and so forth, but it is not up to one what, having done all that, one will experience. This minimal point is what I am insisting on' (McDowell 1996, 10). McDowell has an important point here: it would be implausible to depict a person's experience as so perennially active that the person could always autonomously determine what she experiences, as if did not matter *that* or *how* the world was there. But do we need to put things so starkly, i.e. do we need to oppose passivity and activity in this way? Do we need to say that we must choose either between autonomously determining content and being 'saddled' with it? The answer we want to give here is that we need not choose between these options – there is an option in between (this is what will be called, in Part VI, cognition from the second-person perspective, i.e. a partnership between attentive silence and appreciation), but this option is not easy to see if we continue to think that we can, or need to, silence the sceptic.

The idea of picturing persons as being 'saddled' with content is anything but minimal. McDowell assumes that when we experience, our experience must have a content that is determined by the concepts that person has, such that the person is saddled with the content that those concepts give her. One way to unravel this idea is to ask, as we have been above, why / when do we need to think of experience as having the structure? Why / when would we want to think of experience as broken into one rule-determined-content after another? We have been answering this by saying that we picture experience in this way when we want to silence, once and for all, the sceptic (once one sees this one can then unravel the idea by saying the sceptic need not be silenced). Another way to unravel the idea is to try to make room for an alternative, i.e. to try and show that there is a minimal sense of experience, but it is not one that can be captured by thinking of experience as broken down into distinct contents (let alone rule-determined ones). On such a view, one can *sometimes* think of experience as broken down into contents (one after the other), but this is something one does for certain purposes at certain times – it is a way of characterising or articulating experience that can help us in all kinds of ways, but it does not 'own' experience.

The alternative becomes more plausible as soon as one sees the variety of ways in which one can characterise / articulate 'what' one has experienced: e.g. 'I was swinging on a swing feeling the wind through my hair' is one possible

characterisation, and another is that 'I was at a noisy place with lots of other children screaming and running about', and so on (one could go on here with alternative characterisations / articulations).

One of the confusions that can make us dismiss this alternative is to say that there is always a 'what', an underlying content, which creates the very possibility for alternative characterisations – so that what is offered above as characterisations are but interpretations, and the thesis that experience is conceptual does not mean that there is no room for alternative interpretations (instead, what it means is that the underlying content is determined by concepts). This can be unravelled when one sees that it is only in the course of considering alternative interpretations that one is treating something as an underlying 'what', i.e. as an incontestable specification of content; but that very same thing, when not being used to tie together (to make commensurable) two or more interpretations, is contestable. Thus, it is true that in counting the above two characterisations as interpretations, we can treat as incontestable some understanding of an underlying experience – e.g. a trip with my father and mother to a nearby park – but it is only in the course of so treating the two characterisations as commensurable interpretations that I do not question the 'underlying experience'. I can question it if I say, for instance, I spent my childhood in Vienna where all the playgrounds were sand rather than grass (here, what I took to be the basis for comparing what I characterised as characterisations becomes irrelevant).

What we are trying to show here is how one can be beguiled into thinking that we experience something that has an incontestable content: this happens when we freeze experience and refuse to think it could be anything but what it was. This freezing is a useful tool: it can help us treat, and also evaluate, two statements as interpretations of the same thing (the 'same thing' is precisely the 'what' we are treating as unchangeable for the purpose of comparative evaluation). But the freezing is just that: an instrument we use for certain purposes, and not something that can count as a general theory of experience (i.e. that experience is one such 'what', one such 'freeze frame' after another).

The point in the context of McDowell's discussion of experience as passive (his allegedly 'minimal' sense) is to say that he is right that there is a minimal sense of experience, but that this experience does not have a set content. This is important, for as we loosen the necessity for experience to have a set content, we also loosen the

need for that content to be determined in some way, e.g. by the application of general rules. It is, then, in the context of justification (when we want to make some justifiable assertion about what we experienced) or in the context of systematisation (when we claim there are two different ways of interpreting the same underlying experience) that we think the minimal sense of experience is one that requires a fixed content.

Here, the following reply could be made: but are you suggesting that anything goes – that, for instance, anything could be an interpretation, or that we have no way of evaluating whether some concepts are more applicable or less applicable to some situation (and if you are not, do you not need some notion of 'the situation', which just means you need some minimal sense of 'content')? The question is difficult, but only because it is misguided (basically, it takes for granted the position that is up for debate). The question forgets about the ways in which 'the what', or 'the situation' does not have to be thought of in the way it might be thought of at any one time, and it also forgets that we may not always be interested in comparing interpretations or defending / criticising the applicability of concepts. Thus, it may be true that for the purpose of comparing interpretations (which one is more accurate, say) we may treat a certain 'what' or a certain understanding of 'the situation' as the 'true' one, or the 'real' one, or the incontestable one (we may 'hold it in place' as it were, or treat it as unchangeable), but the point is that this is just an attitude – it does not bestow any special metaphysical status on that which is so treated. This treating is itself active, but it is just one way of being active: another way is to experience the environment as indeterminate, precisely not having (and not needing to be attributed with) a content, or having several different possible contents where the choice as to which one is left open as one continues to relate with the environment.

How, you might ask, does the above make sense of the notion of appearance – of something appearing to one – when it looks as though we can only contrast appearance with reality, i.e. contrast how things appear with how they really are? Here, again, we would say this: we can only ever contrast appearance with appearance, how things appear with another version of how things appear, but what we can do is treat one of those ways of appearing as if it were 'reality', as if it were the standard we use (and one can imagine agreeing with someone on this) for the purpose of evaluating such-and-such a variety of versions of how things appear. By no means does this mean we are somehow arbitrarily autonomous vis-à-vis the

environment: on the contrary, the whole point of allowing for this alternative (of experiencing the environment as indeterminate) is to allow for the environment to appear to us in ways we would not have anticipated, and thus precisely not be in control of. In fact, then, this could be a more positive, more active way of taking what McDowell wanted to take seriously, namely the idea that a person is not always in control of how they experience. Instead of taking this seriously by saddling the person with content, we allow for the person to be surprised by the environment – to be taken in a direction they could not have imagined prior to the relation. The point is to depict the environment as something with which we relate, and thus something that can present itself to us in novel ways – precisely as something different and not familiar. We can relate with the environment in such a way that it can teach us; not something that we immediately characterise or categorise as a result of the concepts we have. In that respect, the opposition between appearance and reality is too stark, and talk (in response to that opposition) of reality (as we have above) as but another appearance that can be treated as a reality can also be misleading (though it does help to some extent, it is a kind of first step): the trick, then, may be to allow for reality, but in a way that can be non-intelligible to us (or better, in a way that need not be immediately intelligible to us).

It may be that what when McDowell appeals to 'responsible freedom' (as he does throughout, e.g. McDowell 1996, 12) he is making a plea for taking a certain (communal) attitude, i.e. of commitment to treating certain things as incontestable for certain purposes (precisely those things that others are also being asked to treat as incontestable for those purposes). Insofar as this would be McDowell's point, we would not disagree with it, as long as it is acknowledged that this 'treating' is an active process, and not the default one that is always and necessarily guiding us until and unless we are given reason to change it. But this, of course, is not McDowell's point: McDowell puts things the other way round, i.e. we have a certain range of concepts, identified by rules, and it is these that make the experience of the world possible, experience always having a content that is justifiable precisely by reference to those rules.

How does McDowell deal with the problem of change? Does he allow for the rule-structured concepts be otherwise than he alleges they are? McDowell does allow for change, but of a very carefully controlled kind. Consider the following:

For example, consider judgements of colour. These judgements involve a range of conceptual capacities that are as thinly integrated into understanding of the world as any. Even so, no one would count as making even a directly observational judgement of colour except against a background sufficient to ensure that she understands colours as potential properties of things. The ability to produce 'correct' colour words in response to inputs to the visual system (an ability possessed, I believe, by some parrots) does not display possession of the relevant concepts if the subject has no comprehension of, for instance, the idea that these responses reflect a sensitivity to a kind of state of affairs in the world, something that can obtain anyway, independently of these perturbations in her stream of consciousness. The necessary background understanding includes, for instance, the concept of visible surfaces of objects and the concept of suitable conditions for telling what colour something is by looking at it. (McDowell 1996, 12)

Notice that we have concepts at two levels here (or, more accurately, we only have concepts properly speaking if we have concepts not only at the first level, but also at the second level): first, the level of getting the colours right; and second, a higher order level that determines whether 'the getting the colours right' is a matter of acting in accordance with standards (rather than just responding to stimuli). There is, says McDowell, 'a network of capacities for active thought, a network that rationally governs comprehension-seeking responses to the impacts of the world on sensibility' (McDowell 1996, 12). Change is possible – 'there must', says McDowell, 'be a standing willingness to refashion concepts and conceptions if that is what reflection recommends' (McDowell 1996, 12-3) – but this change itself occurs within the network (or system), e.g. it is only via the work of concepts at the higher level that we could change our concept of a colour more fine-grained (e.g. split yellow into yellow-with-a-tiny-bit-of-orange and yellow-without-a-tiny-bit-of-orange). The change needs to be justified / justifiable in some way – it cannot be arbitrary or accidental (it must always be the product of the pre-existing order).

The way the entire structure seems to work, then, is as follows: there is the first circle of concepts that is stable – e.g. imagine a tablet of colours, and objects that correspond to them, and then imagining testing someone (or a parrot) for choosing (or pecking) at the 'correct' blob on the tablet of colours when showing a certain object. Here, the possibility of correctness being determined / determinable is set up by holding still (treating as unchangeable) the range of possible colours as standards (in the tablet) as well as the range of objects with possible colours (correctness / incorrectness of the right kind only applies to persons and not parrots, for only

persons are governed by the second circle of concepts). Change is possible, but not within this first circle, and not while evaluating the person (parrot) – rather, we may, after reflection, decide to change the tablet by changing the concept of the colour yellow, but then we need to do so seriously and responsibly (we announce, as it were, the change, so it is not felt to be arbitrary by those are evaluated as correct or incorrect in accordance with it).

The following kind of scenario is left out of this set-up: imagine a droplet of rain falls on our tablet, creating a change in how yellow appears to us – a way of yellow appearing to us that we had not previously experienced. McDowell seems to want to put all the pressure here on a reflective decision: do we express annoyance at the rain, cleaning up our tablet, and restoring it to what it was, or do we (thereby revealing 'willingness to refashion our concepts') take the time to consider whether we should create a different tablet, one more fine-grained, say, which classifies the colours (especially yellow) differently, or perhaps one that elevates yellow to a more central position (the point being we would not have thought it could have such a position prior to seeing how – as a result of the droplet falling – it can be transformed into other colours)? McDowell's identification of this kind of willingness is important, and it is important also to see that (at least on our articulation of the case) there is no determinacy – we could have taken the first option, and we need not have exercised that creative effort of changing the tablet. But there is something that is being missed here: precisely our ability to see (or, better, potentially seeing) how the droplet falling might give us reason to change the tablet. In other words, there is room for the description of an experience of relating with an environment that surprises you: you see yellow in a way you could not have been prepared for; your experience of yellow is not a product of your prior conceptual scheme. Surely, putting things this way, one can see that one cannot even exercise 'willingness to refashion concepts' without this ability to relate with the environment where how the environment ought to be understood is being explored (and not immediately available via pre-existing concepts)?

When we picture cognition as necessarily conceptual, we picture the mind as too eager for reducing whatever it experiences to how it has previously classified things as appearing to it, or we put all the explanatory pressure on moments of reflection where we re-establish, stabilise again, a conceptual scheme (neglecting to see what is prior to this moment of reflection). As a result, the mind is pictured as

either too willing to be comfortable and surround itself with familiarity, or as too reflectively careful and serious, hardly capable of surprise, always controlling what change (in itself) it might endorse. Another way to put this is that we lose too much of the mind when we think of it as constantly making justified / justifiable judgements, whether in support of a claim as to how things appear to it or in support of a change as to how things ought to appear to it in the future.

When McDowell pictures the mind as governed by higher-order concepts (and in so doing attempts to distinguish the mind of a human being from the parrot) – e.g. by the concept of visible surfaces of objects – he is failing to imagine circumstances where we can be responsive to colour without such concepts mediating us. It is when we fail to imagine alternatives – or when for some reason we want to hold those higher-order concepts in place, treat them as unchangeable standards – that we think our cognition is necessarily mediated by concepts. Imagining alternatives takes effort, but it is possible: for instance, imagine experiencing the colour of the inside of an apple's skin or imagine experiencing the colour of a sunset (the point being that these are examples of experiencing colour where your experience is not mediated by the allegedly second-order concept of something appearing as a colour only when it is a visible surface of an object).

McDowell does not consider such possibilities because his argument has a different structure, i.e. it is based on the idea of setting up and defending a certain limited range of conditions of possibility – the picture of thinking, for instance is set up by articulating what is thinkable (see, e.g. McDowell 1996, 28). The general problem with 'conditions of possibility' argumentative structures is that they describe cases in a way where it is difficult to see how that which is described as having been experienced could have been experienced in any other way, for what is described as having being experienced is said to have been possible only because of certain pre-existing conditions. This is another instance where the explanation – rather than description – drives the example. With such cases, one needs to crack them open, and show that the conditions of possibility could never possibly anticipate all of what may be experienced. The way to do this is to prioritise description, but do so in a way that conveys the sense in which what may come to be experience is not pre-determined. We have been performing this task throughout this thesis, as many of the examples used to buttress the governance view need this kind of dismantling.

In chapter 3 of *Mind and World*, McDowell tackles the idea of 'non-conceptual content', and thus raises issues of relevance to the ways in which we have been arguing against necessarily conceptual cognition. We need to take a look at these arguments before turning to the debate between Dreyfus and McDowell. McDowell's thesis, to recall, is that 'we must insist that the understanding is already inextricably implication in the deliverances of sensibility themselves. Experiences are impressions made by the world on our senses, products of receptivity; but those impressions themselves already have conceptual content' (McDowell 1996, 47). The objection (i.e. the argument for the possibility of non-conceptual content) fastened on by McDowell (in reply to that thesis) comes in the form of a claim made by Gareth Evans (from his *Varieties of Reference*, 1982). According to McDowell, Evans argues that although judgements necessarily involve conceptualisation, the experiences that judgements are based on do not (McDowell 1996, 47). Evans refers to these 'experiences' as 'informational states which a person acquires through perception' (Evans, quoted in McDowell 1996, 47), and he argues (in McDowell's summary) that these states are part of an 'informational system' which is 'the system of capacities we exercise when we gather information about the world by using our senses (perception), receive information from others in communication (testimony), and retain information through time (memory)' (McDowell 1996, 48). The point is that for Evans these informational states (which we are said to share with animals) are non-conceptual. This is how McDowell articulates the difference between his view and Evans's view:

> According to Evans, conceptual capacities are first brought into operation only when one makes a judgement of experience, and at that point a different species of content comes into play. Contrast the account I have been urging. According to the picture I have been recommending, the content of a perceptual experience is already conceptual. A judgement of experience does not introduce a new kind of content, but simply endorses the conceptual content, or some of it, that is already possess by the experience on which it is grounded. (McDowell 1996, 49)

We shall look at McDowell's arguments in a moment, but let us pause for a moment to consider this way of arguing against the necessity of non-conceptual content. Recall what was said above about the kind of minimal notion of experience we should make sense of: it was said that this could not be a notion that could be accompanied –

or explained – via the idea of a fixed content. That Evans is working with this notion of fixed content is clear from the way he speaks of experience, e.g. as an 'informational state', as something that can serve as 'the basis' for judgement. 'Informational state' and 'basis' are inadequate notions for the minimal sense of experience we need (or so it has been argued above) to defeat the argument for the necessity of conceptual content. If the minimal sense of experience is to work, it must be possible to experience things in a way that offers contingent opportunities – invitations as it were – for specification, i.e. for constructing bases or informational states if the need arises (it is because one *relates with* the environment that such opportunities or invitations – which must be able to surprise one – become possible); the minimal sense of experience, as we wish to make room for it (in order to defeat the thesis concerning the necessity for conceptual content) is defined out of existence if what is said to be foundational is an 'informational state' or a 'basis'.

The same notions, then, that McDowell uses are in fact used by Evans: the notion of a system, of a ground or basis, and of a state or content, with the difference that Evans just applies them at the level of the senses (we shall put aside issues concerning testimony and memory). We need, however, more tentative notions (such as the notions of opportunity and invitation) for a non-conceptual understanding of the senses learning and working. This does not mean that we must say that every time we experience, we only ever experience in this minimal sense, i.e. that concepts play only a backward looking role. This would be going too far, for it is possible for our senses to be primed in such a way that they experience what our use of the concept directs them towards (this is a case of artificially determining relevance),[122] but we make a mistake when we think that we must by necessity always direct our senses via concepts (as if what our senses might detect as relevant was always determined by defeasible default).

So long as the debate between conceptual and non-conceptual content is set up in such a way that it divides matters into either concepts always necessarily guiding *ex ante* (McDowell) or them always and necessarily serving *post facto* (Evans), where this division itself is based on common commitment to concepts such as 'basis' or

---

[122] We shall see later (in Section 15.A.) that social psychologists have used this idea to argue for a kind of manipulation of the environment to make certain 'norms salient'. As is argued in that Section, this is possible, but we should not generalise it as a description of how the mind, by default, learns and works (i.e. as if it but learnt such norms, and was thereafter perennially primed to see what they, on this view, determine ought to be seen).

'informational state', then we will not defeat the thesis that cognition is necessarily conceptual. What we need is to avoid a view of the debate in an all-or-nothing manner: it is possible for us to (on occasion, and in an active way) rely on concepts; it is also possible for concepts to play what McDowell refers to as an 'actively self-critical' role, or even 'control' as he puts it (McDowell 1996, 49) – i.e. we can still discipline ourselves actively and self-critically in accordance with some normative resources that we treat (for that purpose, and thus during this process) as unchangeable, but this does not mean that they are unchangeable in principle, or that they control us in such a way that we can only ever experience that which they make possible to experience.

Let us return now (keeping the above provisos in mind) to McDowell's argument against Evans. McDowell's reply is marked by certain anxieties: first, the anxiety that 'we would not' (if we accepted Evans's picture) 'be committed to having different stories to tell about the sentient lives of rational and non-rational animals' (the demarcation problem); and second, the anxiety that without concepts 'empirical content threatens to degenerate into a picture of a frictionless spinning in a void' (again, notice the all-or-nothing nature of this anxiety; McDowell 1996, 50). Both anxieties lead McDowell to posit 'external constraints', but not external constraints that 'could at best yield exculpations where we needed justifications' (McDowell 1996, 51), but rather external constraints of the right kind. Thus, the question that he puts to Evans is whether Evans has just such external constraints of the right kind. The answer is obvious: Evans does not have such constraints (in a way, Evans tries to show us how we might experience without such constraints (ever) leading us – but still possibly serving as standards for evaluation – but, as argued above, his account suffers from the fact that he tries to force this notion of experience into having a 'content'). McDowell, in short, takes Evans to offer a picture of 'intuitions' without concepts, which as McDowell reminds us, Kant long ago told us were 'blind' – Evans, then, according to McDowell, offers 'a version of the Myth of the Given' (McDowell 1996, 51).

Our answer to this critique by McDowell of Evans is as follows: Evans opens himself up to this charge precisely because the disagreement about experience is being fought on a stage where what matters is how we might be justified in taking what appears to us as what there really is to experience (the notions of justification and content go hand-in-hand on this platform). Consider the following: 'I am

claiming', says McDowell, 'that although Evans does take care to credit experiences with content' (that McDowell sees this as a positive thing should already ring bells for us) 'that does not save them from being intuitions in a sense that entitles us to apply the Kantian tag to them: since they are without concepts, they are blind' (McDowell 1996, 54). It is, then, the postulation of content that puts Evans within the scope of McDowell's argument (and the reason, it seems, that Evans postulates content is because he still wants experience to play the role of a basis, of some ground for judgement – you could say that Evans splits justification into experiential content and the articulation of judgement, the point here being that this still prioritises the context of justification).

Notice how this plays itself out in the context of our experience of colour. Evans, says McDowell, is impressed with how fine-grained our experience of colour is, and he thinks it unlikely that we must have concepts guiding us every time we have an experience (with content) of a particular shade. McDowell, in reply, says that we should not hold on to a stock of expressions (e.g. 'red', 'green', 'burnt sienna'), for as soon as we notice that we must be experiencing something specific, 'that shade', we also (ought to) realise that 'that shade' has to be a concept: as he puts it, '"that shade" can give expression to a concept of a shade; what ensures that it is a concept – what ensures that thoughts that exploit it have the necessary distance from what would determine them to be true – is that the associated capacity can persist into the future, if only for a short time, and that, having persisted, it can be used also in thoughts about what is by then the past, if only the recent past' (McDowell 1996, 57). He then adds, 'What is in play here is a recognitional capacity, possibly quite short-lived, that sets in with the experience. It is the conceptual content of such a recognitional capacity that can be made explicit with the help of a sample', which then 'Later in the life of the capacity…can be given linguistic expression' (McDowell 1996, 57-58).

We see several things going on here: first, notice that what is working in the background to produce this image of the mind learning and working is the need for 'thoughts…to have the necessary distance from what would determine them to be true'; second, notice that the effect this need has is one where the mind is pictured as only ever being capable of experiencing that which it 'recognises' (as if the tablet of possible shades was always and already inscribed in it, but it had not been aware of it – now it can make the shade explicit, and also give it a name). To unravel this – as we have been trying to do – one needs first to see that thoughts can be thoughts without

there always being an external constraint of the right kind (i.e. the kind that can give us the 'reassurance' McDowell is after – not mere exculpation, but instead justification), and we have tried to make room for those kinds of thoughts by speaking of minimal experience (i.e. experience as an opportunity or invitation by an environment one is relating with) and also how the mind appears to us when we prioritise the context of appreciation (as opposed to justification); then, second, one needs to see that the mind can learn and work in ways that cannot be captured by the notion of recognition, i.e. that the mind can learn things that it does not always and already recognise (although, of course, this does not mean that once the mind has learnt something, it cannot go on to recognise it).

We need to say a bit more about this second observation: first, it is possible for us to learn something such that we recognise it (although even here 'recognition' is an evaluation of sufficient likeness, it is just that our capacity is not accompanied by awareness that something is sufficiently alike); second, it is possible for us to learn something such that it can help us form an ability to see other things as sufficiently like it (here, this is something more actively creative than the first, for although both depend on an evaluation of sufficient likeness, this extends the likeness in greater degrees); and third, we can have experiences (including ones from which we learn) that do not have content but rather offer opportunities for articulation / specification (now, or later – and later in a way that could change the way we articulate / specify them now).

McDowell, it seems, would not accept the second two, and he would query the characterisation of the first: it is, he would say, all recognitional ('recognition all the way down'). One could try to argue against McDowell (as Evans does) by accepting the recognitional story, i.e. accepting that we must have experiences with content that do not require any evaluation of likeness, but then one opens oneself up (as Evans does) to precisely the argument thrown up against him by McDowell (that he just gives us (impossible) instances of blind intuitions). That is why we have been urging here that we need to not accept the recognitional story, and thus not accept that we must have experiences with content. We have argued that this does not mean that we cannot, when experiencing, treat something as a standard according to which we evaluate ourselves, and nor does it mean that we cannot exercise a recognitional capacity (though even this is more creative – i.e. requiring evaluations of likeness – than we might at first think). What it does mean is that we can have experiences and

appreciate things in ways that cannot be forced into the mould of making judgements on certain grounds, or justifying ourselves, or being reassured (i.e. having a ground we might potentially refer to if challenged) even when we do not want to be reassured (as the idea of 'that shade' being a concept suggests we can be).

What we are seeing here is how the idea that we must be able to say whether we are correct or not is linked to the positing of external constraints as somehow always and necessarily present (even before we know it – they are there implicitly, waiting to be made explicit) and determining what is correct or not (whether what appears to us as that shade really is that shade), and how this is linked to a picture of the mind as always and necessarily guided by those constraints, with the result that we radically narrow our understanding of how the mind can learn and work. Put in the context of our experience of colour: we can see that we can have experiences of colours that we have not recognised before (in a way that McDowell does not) as long as we make room (again, in a way McDowell does not) for picturing the mind either as having minimal experiences or as exercising capacities in the context of appreciation; put more negatively, it is only under the guise of the context of justification (narrowly understood as requiring but mechanical recognition, and no evaluation of likeness) that we cannot but experience some specific colour as a shade, 'that shade', where our capacity to recognise it is enabled by a concept that we can then make explicit (by creating a sample – and introducing it into our tablet) and giving it a name.

The above argument by McDowell in the context of colour experience is in fact more ambiguous than we have presented it. The ambiguities step in when we try to understand in what sense exactly 'that shade' is a concept. We said above that McDowell says that '"that shade" can give expression to a concept of a shade' (McDowell 1996, 57). This term 'can' is ambiguous, for in fact McDowell does not want to say it may not be a concept: he wants to say it has to be (recall that he says that 'that shade' is already a concept, not just that it can lead to one). On McDowell's explanation, we cannot experience a colour if it is not 'that shade', for it being experiencable means it must be a concept. But McDowell also realises that this colour we confront might simply be forgotten by us, and never transformed into a concept. Seeing this, McDowell presents this experience as a kind of concept-in-waiting, but from our perspective the key question is: why must this 'concept-in-waiting' (something that could be transformed into a concept) itself be a concept? Here is

where McDowell is very difficult to pin down. Recall the following: 'what ensures that it [i.e. that shade] is a concept…is that the associated capacity can persist into the future, if only for a short time, and that, having persisted, it can be used also in thoughts about what is by then the past, if only the recent past' (McDowell 1996, 57). But why should something that is capable of being transformed into a concept, i.e. on this view, as something that gives us a recognitional capacity in the future, be itself – in this primordial state – a concept? Why does McDowell think this 'persistence into the future', even if 'short-lived', is enough to make it into a concept (as well as, potentially, a full-grown concept, so to speak)? This is very difficult to answer: perhaps McDowell is attributing (in a way that does not seem warranted) some autonomy to this concept-in-waiting, as if it could enable recognition just like that, with no effort required to see that something is sufficiently alike for it to be treated as an extension of that concept-in-waiting. In other words, perhaps if one sees that an active evaluation of sufficient likeness is required to extend the concept-in-waiting, then one also sees that 'that shade' is not a self-contained, autonomous package, that applies (persists into the future) automatically. Perhaps if one then sees that, then one would be more willing to say that it is not a concept. In any event, the burden is here on McDowell to tell us why this 'persistence into the future', even if 'short-lived', combined with this potential transformation of 'that shade' into a concept, is enough to make it itself a concept.

Another way for McDowell to argue for 'that shade' being a concept is to say that we could not see it unless we have the concept for 'shade', such that the only reason we can see specific shades (that, and that, and that shade, etc) is that these are, as it were, applications of the higher-order concept of a shade; seeing such-and-such specific shades is dependent on the higher-order concept of a shade (McDowell 1996, 58). But this too is problematic: first, even if one grants McDowell the idea that we have a concept for 'shade', it is not clear why this means that every time we see a specific shade we recognise a concept (i.e. it is not clear why, even if it were true that every time we experience a colour we apply concepts one of which is the concept of a shade, that this would mean that what the concepts are being applied to must themselves be concepts);[123] second, we should not grant to McDowell the idea that we

---

[123] We leave aside the question of whether McDowell's view implies that our experience would need to be accompanied by an awareness that we are seeing 'a shade', for if it did, this would surely be

have a concept for 'shade', for what we might think of as a shade is not somehow always and already defined or determined by the nature of 'shade' (in other words, the very division into first-order and second-order concepts is misleading, for it makes us neglect to see the indeterminacy of second-order concepts; the point is tricky because when it is treated, in any instance, as a second-order concept, then it is being treated as if it were determined, our point being that that treating does not give it any metaphysical status).[124]

It is hoped that we have given something of the flavour of McDowell's approach in *Mind and World*, and also sufficient argument to suggest that there is room for resisting the way in which, in that book, he argues for cognition being necessary conceptual (or, put differently, for experience of the world being necessarily mediated by concepts). We have also seen that there are ways of arguing against McDowell (positing non-conceptual content) in a way that does not help (but just digs us deeper into the picture we need to resist).[125] To help us get clear on how certain objections to McDowell's view do not help us, but also to get a better understanding of McDowell's more recent views, let us briefly turn to an instructive recent debate between McDowell and Hubert Dreyfus on the pages of the journal, *Inquiry*.

B. Dreyfus vs. McDowell

The debate between McDowell and Dreyfus really begins with Dreyfus's Presidential Address to the Pacific Division Meeting of the American Philosophical association in 2005, so that is where we must look first.

In this 2005 Presidential Address, entitled 'Overcoming the Myth of the Mental' (2006), Dreyfus argues that 'philosophers who want to understand knowledge and action can profit from a phenomenological analysis of the non-conceptual embodied coping skills we share with animals and infants' (Dreyfus 2006, 43), and in so arguing, he targets McDowell's thesis concerning the conceptual nature of human

---

implausible: surely we do not, and we do not need to, say to ourselves: 'ah, this is a shade (of such-and-such a colour) I am seeing'.

[124] The explanatory burden can be put on McDowell in the following way: we can ask him what are the criteria that make a shade a shade, and then give him a counter-example, i.e. of how it is possible to experience something as a shade in a way that is not an application of the criteria he announces. The indeterminacy of the concept cannot be hidden under the carpet of a potential second-order treating.

[125] To be fair let us repeat that we have not here taken Evans on his own terms, but have referred to McDowell's reading of Evans – and this is because our focus is McDowell in this Chapter.

cognition. It is not always clear what Dreyfus is trying to explain: is it knowledge? Or is it action? Or is it 'intelligence' (Dreyfus 2006, 43)? These explanatory targets differ. Further, what explanatory target one has will have an effect on how one pictures the mind learning and working. Quite often, in these debates (over whether cognition is conceptual or not), the context is epistemological, and thus the interest is in explaining knowledge. This already makes it more likely that the context of justification – or, other ideas, such as the search for a ground or basis, or the search for something anterior to explain what comes after, and so on – will be prioritised, and we have already seen what effect that has on our picture of the mind.[126]

One of the ways in which Dreyfus sets up the dilemma discussed in his paper is by referring to the frame problem. We encountered the frame problem in the previous Chapter (see Section 11.C.). As Dreyfus reminds us, one of the great problems for the computational model of artificial intelligence is the problem of relevance (or put differently, if one is thinking in terms of rules, the applicability of the rules). As was observed above, there are different ways one can read the frame problem, and how one reads it has an important effect on what one takes to be a proper answer to it. One way to read it is to ask, 'what determines what is relevant'? In other words, one way to read the frame problem is to read it as a problem about what determines what is relevant in some situation. Another way of reading the frame problem is that it reveals the very limits of thinking that relevance can be determined, i.e. that it is a reminder that relevance cannot be determined (here, to recall, the frame problem is not so much something that can be solved, but a warning against the ambition of trying to determine relevance). Although again matters are not always very clear, it does seem to be the case that Dreyfus reads the problem in the first way (i.e. he tries to solve the problem of how relevance is determined). Consider, for instance, the following question: 'How, then, do we manage to organize the vast array of facts that supposedly make up commonsense knowledge so that we can retrieve just those facts that are relevant in the current situation?' (Dreyfus 2006, 44). Here, the phrase 'just those facts that are relevant in the current situation' suggests that Dreyfus thinks that what is relevant is determined somehow. This also appears to be

---

[126] Another way of putting this is to say that our experience can be valuable to us, can be rich and informative, without being knowledge (at least where 'knowledge' is understood to be necessarily subject to the context of justification). Perhaps, rather than speaking of 'knowledge' it is better to speak of learning.

the reading we would get from the following passage (which answers the above question):

> We can't manage it any more that a computer can, but fortunately we don't have to. Only if we stand back from our engaged situation in the world and represent things from a detached theoretical perspective do we confront the frame problem. That is, if you strip away relevance and start with context-free facts, you can't get relevance back. Happily, however, we are, as Martin Heidegger and Maurice Merleau-Ponty put it, always already in a world that is laid out in terms of our bodies and interests, and thus permeated by relevance. (Dreyfus 2006, 44)

Here, Dreyfus again seems to assume that what is relevant is there for us to recognise as relevant; we are passive vis-à-vis relevance. What appears to interest Dreyfus, then, is how we access what is relevant. That access is not one that can be modelled on the computer model, but rather is embedded in 'a world that is laid out in terms of our bodies and interests, and thus permeated by relevance' (Dreyfus 2006, 44). Here, the image of a world being laid out again suggests that what is relevant is there, somehow, in the situation – and the only question is how we access it (via our bodies – via some notion of absorbed embodied coping).

It was mentioned above that matters are not always clear in Dreyfus. Thus, for instance, Dreyfus contrasts his account with an account that would wish to see us as merely responding to 'fixed features of the environment', somewhat like robots who 'don't feed back into their world what they have learned by acting in it' (Dreyfus 2006, 44). There is ambiguity even within this one phrase, for why say 'learned' rather than 'learn', i.e. why not allow for learning on the run what might be relevant? The reason is that Dreyfus is not here distinguishing himself from a position where we acknowledge the limits of an approach that places (and describes the actions, intelligence, etc, of) human beings in contexts where relevance is not an issue, but rather from a position where it is thought that just because what is relevant is fixed (or not questioned) means that what we respond to in the environment is fixed. Dreyfus says: 'We need to consider the possibility that embodied beings like us take as input energy from the physical universe and process it in such a way as to open them to a world organised in terms of their needs, interests, and bodily capacities' (Dreyfus 2006, 45). Here, Dreyfus uses the image of a 'world' that is already 'organised in terms of' our 'needs, interests, and bodily capacities.' This suggests precisely a world

in which what is relevant is fixed. What we could, generously, take Dreyfus to be saying is that when we are 'open to the world' organised in terms of our needs, interests, etc, our needs, interests etc are not such that they determine what we respond to in the environment. However, although leaving open 'what' we may respond to in the environment is an important step (and from the perspective of this thesis, a positive one), Dreyfus nevertheless positions us as choosing between things that have always and already been structured (by our interests, needs, etc). The pool of what could be relevant is pre-determined. Dreyfus does position the mind somewhat actively, but he still subjects the description of the mind learning and working to how it is situated in, and how it copes with being situated in, a world that is always and already laid out or organised in some manner (e.g. by our interests, needs, etc).

How does all this affect the way in which Dreyfus criticizes McDowell? It means, in short, that Dreyfus cannot get at the heart of McDowell's approach. Essentially, it means that Dreyfus tackles McDowell at the level of accessing what is already imagined by the theorist to be relevant. Thus, for instance, Dreyfus criticizes McDowell for an understanding of *phronesis*, or ethical expertise (on Dreyfus's translation), as necessarily 'conceptually articulated', arguing that 'On the contrary, *phronesis* shows that socialization can produce the kind of master whose actions are so context dependent that he could not rely on habits based on reasons to guide him' (Dreyfus 2006, 46). Dreyfus positions his attack at the level of a person's access to relevance always being guided by reasons (consciously or habitually), but he only replaces guidance by reasons with guidance by 'socialisation' – in both cases, persons are depicted as having to access what is always and already laid out and organized, or always and already defined (though not necessarily by a rule or a reason) as relevant.

It is important to see that Dreyfus's position here is not one that argues that once we learn what is relevant by learning the rules (which we at first reason with deliberately) we then come to process the rules sub-personally (unconsciously or implicitly). That is not Dreyfus's position. Indeed, he criticizes this position very colourfully, by saying that to 'To assume that the rules we once consciously followed become unconscious is like assuming that, when we finally learn to ride a bike, the training wheels that were required for us to be able to ride in the first place must have become invisible' (Dreyfus 2006, 46). This is important, because it makes available a description of skills that are not a matter of simply subpersonally or unconsciously

following rules (i.e. a matter of developing a habit to follow rules). But the problem is that it does not get at the root of the issue, for although there is a difference between what Dreyfus refers to as 'analytic rule-following' (either personal or sub-personal) and 'engaged, holistic mode of experience' (Dreyfus 2006, 47), the latter still positions a person (or an 'expert') within a situation that is somehow already laid out or organized. One sees this at work in Dreyfus's discussion of expertise, and his drawing on the example of chess: Dreyfus makes room for learning to make ever-finer discriminations, for learning to distinguish between different situations requiring different reactions – thereby developing 'the immediate intuitive situational response that is characteristic of [chess] expertise' (Dreyfus 2006, 47) – but he sees all this as happening: 1) generally, within a structure, i.e. of a game of chess, a game which he understands as nevertheless having a certain nature (constituted by certain rules) into which persons are socialized; 2) more specifically, with respect to any situation (in chess), as one in which what ought to be done is laid out and the key question is the kind of skill (corresponding to the level of expertise) we might use to do what we ought to do (make the correct move).

This picture of expertise should be a familiar one by now: the novice uses rules, e.g. about strategy and tactical computation, to reason his way to a correct solution or to remembering the pieces on the board, whereas the expert has a 'immediate intuitive situational response'. What Dreyfus is opening up is the possibility of understanding expertise in a way that is not driven (consciously or not, deliberatively or not) by rules (so there is no question here of how rules are internalised, because Dreyfus's point is that they do not have to be), but what he is closing is the possibility of understanding the mind learning and working in any other way than in situations where what is relevant / correct is fixed (determined in Dreyfus's account, by our needs and interests). This means that we miss the sense in which the mind can create what might be relevant or correct (thereby also potentially creating a need or interest), and missing this means we describe the mind learning and working too passively.

Now you might ask what any of this has to do with McDowell. The question is a good one. For McDowell says nothing that is strictly at odds with much of what Dreyfus says. Dreyfus thinks he is arguing against McDowell because he argues that the expert is not necessarily exercising conceptual capacities when he develops a 'perceptual ability to respond to subtle differences in the appearance of perhaps

hundreds of thousands of situations' (Dreyfus 2006, 47-8), but we have already seen that McDowell allows for this possibility, but just understands our capacity to so discriminate to be a matter of targeting concepts (each shade is a concept) as well as presupposing concepts (e.g. the very notion of a shade). Dreyfus does not get at the root of the problem with McDowell's account because he accepts the basic picture of the mind confronting what is there to be recognised (and 'possessing', as it were, one content – one shade – after another). Certainly, Dreyfus is knocking on the door by speaking of 'hundreds of thousands of situations', but the problem is that the issue is not a matter of numbers, but of where one positions the mind (when accounting for how it learns and works), i.e. whether one positions it (as both Dreyfus and McDowell do) in the passive position of having to recognise / respond to what is assumed to be there (the differences in those situations, or that array of shades, lying hidden, and awaiting being made explicit), or whether one positions the mind where it confronts an open future (e.g. one in which the mind can create what might be relevant).

That McDowell thinks that he is not quite the target of Dreyfus's attack is very clear from his response (2007a). What McDowell says, quite reasonably, is the following: 'Dreyfus thinks the sphere of the conceptual in my sense, the sphere of the rational, cannot include embodied coping. He thinks embodied coping skills are, just as such, non-conceptual… But why should we accept that embodied coping skills are, just as such, non-conceptual?' (McDowell 2007a, 339). And then McDowell adds: 'I do not have to ignore embodied coping; I have to hold that, in mature human beings, embodied coping is permeated with mindedness. And that is exactly what I hold' (McDowell 2007a, 339). What we find here is to two theorists stamping their feet and banging the table: one saying that we should call our ability to discriminate between shades as exercising concepts, and the other as exercising embodied skills – but neither can defeat the other, for both are positioning the mind in the same spot, namely as facing something to be recognised (thereby not allowing the mind to create what can be recognised). McDowell complains that Dreyfus reads his notions of concepts and rationality much too narrowly, and, indeed, he has good grounds for that complaint, for he reads them and expresses them very broadly in *Mind and World*. For instance, he does not read rationality (as Dreyfus accuses him of doing) as situation-independent or detached, but instead describes learning to do and doing what a situation requires in an engaged way as learning to recognise concepts and

presupposing conceptual abilities.[127] One might, of course, argue that McDowell is stretching the notion of concepts and rationality too far – indeed, what emerges in the debate is how differently McDowell and Dreyfus are using these words – but the point is that one does not get at root of the problem for the problem does not lie so much in what we call what needs to be explained (though this too has a certain importance), but more in what we think needs to be explained.

What McDowell can do to reply to Dreyfus – and this is essentially what he does – is either say that Dreyfus ascribes too narrow a reading of concepts and rationality to McDowell, or relegate many of Dreyfus's observations to 'phenomenological' status. That this second strategy plays a large part of McDowell's response is clear from the following: 'Acknowledging the pervasiveness of mind in a distinctively human life is consistent with appreciating those [i.e. Dreyfus's] phenomenological insights' (McDowell 2007a, 346). In other words, Dreyfus has not got at the root of why McDowell thinks we need to depict human life as pervasively minded (or, in other words, necessarily conceptual). Indeed, the way Dreyfus argued against McDowell made him vulnerable to a criticism that he himself was perpetuating a myth – not the Myth of the Mental (as he accused McDowell of perpetuating), but the Myth of the Disembodied Intellect (as McDowell accused him of perpetuating; McDowell 2007a, 349).

Dreyfus has replied to McDowell's (2007a) reply and focused in his reply on the claim of 'pervasive mindedness' (Dreyfus 2007a). He argues, in this new critique, that 'Where I differ from McDowell is that I hold that situation-specific mindedness, far from being a pervasive and essential feature of human being, is the result of a specific transformation of our pervasive mindless coping' (Dreyfus 2007a 353). In so doing, however, Dreyfus offers more of the same kind of criticism. He does so by fastening onto the limits of McDowell's notion of freedom (which we noted briefly above) as stepping back and reflecting. He argues that this is not the only notion of freedom – that apart from this, we also have the freedom 'to let ourselves be involved' (Dreyfus 2007a 355) – the point here being that when we let ourselves be involved, and in actually being involved, we are not acting mindfully or mindedly, but rather mindlessly. This is more of the same because McDowell does not limit

---

[127] Certainly, nowhere does McDowell say that his commitment to seeing such situation-dependent and engagement as conceptual means that he requires deliberation or even inference-making: see, e.g. McDowell 1996, 49, fn. 6; he also says as much in his reply, McDowell 2007a, 340-2.

mindedness (even if he limits freedom) to stepping back; on the contrary, his point is that mindedness is pervasive, and pervades also our engagement in activities.[128] It is also more of the same because it still positions the discussion around 'the situation' or around 'that which is relevant' – just making a distinction between either stepping back (McDowell) or letting oneself be involved (Dreyfus), but not getting at the root of the problem, which is the assumption of 'the situation' or 'that which is relevant'. Matters are also not assisted here by trying to reverse the condition of possibility (as Dreyfus attempts to do, e.g. Dreyfus 2007a, 355-6), i.e. by saying that mindedness is made possible by mindlessness, for McDowell would just rephrase this to read: sometimes, what we phenomenologically experience as 'mindlessness' is indeed operating in the background to enable 'mindedness*' (where mindedness* is understood phenomenologically, i.e. accompanied by deliberation, as when one steps back), but this does not mean that this phenomenological mindlessness is not permeated by mind.

But there are aspects of Dreyfus's reply that do help seeing some of the limits of McDowell's account. One example is where Dreyfus says that the objects of the mind are not 'affordances' as such, but rather the 'solicitations' of those affordances, e.g. not the open mouth of the cave, but the way the cave invites us to come in (see Dreyfus 2007a, 356-60). This is important, and it takes us back to the earlier discussion concerning what one is trying to explain: is it knowledge, or action, or intelligence? For here, as Dreyfus says, when one pictures the mind as responding to solicitations, one resists thinking of the mind as 'focusing on and naming fixed features' (Dreyfus 2007a, 358). Dreyfus is getting closer here to an object of explanation that is more tentative. He points to something being too static, too fixed, about the kinds of objects McDowell puts in front of the mind – what it sees as doing (in McDowell's language, as recognizing). He argues that we cannot think of the mind as detecting 'determinable, nameable and thinkable facts', as if we were only open to 'apperceiving categorically unified facts', or 'propositional structures', and he asks us instead to think of 'indeterminate solicitations to act' (Dreyfus 2007a, 359). One can see Dreyfus groping here for something promising from the perspective of

---

[128] We saw, above, that in *Mind and World* McDowell makes special mention of the fact that he does not understand the passivity of experience to mean absence of active searching, active response to affordances etc; indeed, in his reply to Dreyfus's reply, McDowell 2007b focuses on the way in which his account of responsiveness to reasons is active, not 'detached'; this theme is also the theme of McDowell's manuscript on the philosophy of action, e.g. in it he argues that 'rationality can be in bodily activity as opposed to behind it': McDowell unpublished, 16.

this thesis, but the obstacle he confronts is that he still positions the mind (still understands / describes it by seeing it learning and working) in an environment already laid out and organized, where the solicitations (no matter how 'indeterminate' he labels them) are still depicted as there, needing to be detected – rather than created – by the person). Again, it is important to see the positive side of this: Dreyfus pushes us in the right direction, towards something 'indeterminate' and tentative, and he does so by challenging the nature of the objects that (he takes) McDowell posits for the mind. We will need to build on this when we try to understand / describe the mind without postulating an environment for the mind to act in.[129]

A note needs to be added here about reference, in the above discussion of the debate, to the person 'creating' what might be relevant. Does this emphasis on a person's creativity mean that we are falling back into a kind of arbitrary say-so? The answer is no, and it is important to see why. The possibility of creating relevance is important to stress, for without that idea one all too easily positions persons passively in an environment that is already laid out and organised for them – where relevance is determined for them. But positioning persons actively does not mean losing sight of the environment: on the contrary, the focus of activity is on the quality of the relation between person and environment, i.e. precisely the ability of the person to be surprised by the environment, or be taken in a direction the person would not have anticipated. Relevance, then, is created (potentially) in the relation, and not by the person imposing something on the environment; this does not mean the environment is the 'absolute' or 'ultimate' source, and that this is so because it is possible for the person to not respond to – to not take up the invitation or opportunity issued by – the environment. One might ask about the possibility of surprise, and here two answers can be given: first, someone can be surprised quite by accident; and second, someone can be surprised by looking for surprise – by actively allowing herself to be surprised (thereby, in a way, inviting the environment to invite her). Both ways must be made room for.

We have, of course, not touched here on all aspects of the debate, which not only stretched over several papers (Dreyfus 2006, 2007a and 2007b, and McDowell

---

[129] When McDowell replies to Dreyfus's reply, he asserts that 'the distinction Dreyfus insists on between affordances and solicitations does not amount to much' (McDowell 2007b, 369). One way to say what we arguing above is that we do think it amounts to a lot, but perhaps one can also see it does when one does not position the mind in the way that Dreyfus does, i.e. the power and importance of Dreyfus's distinction is not seen when the mind is positioned as passively having to detect solicitations. This would explain why McDowell does not see the difference.

2007a and 2007b), but which is also the subject of a forthcoming book of papers by other philosophers responding to the debate (see Schear, forthcoming). What we have seen, in briefly examining the debate, is that McDowell's view has not changed in essence, though he has taken the opportunity to clarify his view.[130] More importantly, for present purposes, what we have seen is how (both by reference to Evans and Dreyfus) not to tackle McDowell (through Dreyfus does give us some hints as to where to put more pressure on McDowell's view). We have seen, that to take on McDowell, one needs to dig deep into the motivations for understanding human cognition as necessarily conceptual or, put differently, for presenting experience of the world as necessarily mediated by concepts. We have tried to do this in some detail, unraveling these motivations as best we could, while still trying to work with the kinds of examples that McDowell gives us and offering alternative readings of these examples. In offering alternative readings, we have tried to make room for other possible ways of seeing the mind learning and working, e.g. as in the version of minimal experience we (briefly) proposed and in the idea of the context of appreciation, which has re-appeared throughout this thesis. Later, in Parts V and VI, we will be building on these glimpses of alternatives. For now we hope to have shown that it is only when one sets things up in a certain way, pursuing certain objects of explanation, that one presents cognition as necessary conceptual, and that one does so to the detriment of our understanding / description of how the mind learns and works.

---

[130] We have several times referred to McDowell's unpublished manuscript on the philosophy of action, but this too does not seem to indicate any shift in position – apart from the suggestion that the discontinuity between animals and human beings was perhaps over-emphasised in *Mind and World* – but as this is not yet published and was accessible only in draft form, we will here refrain from analysing it at depth.

Chapter 13. Adequately, Yet Unreflectively

An important recent attempt to carry on the Dreyfus project, though with distinctive contributions of its own, is the work of Erik Rietveld. In a series of papers (2008b, 2008c, 2010, and forthcoming), all of which draw and develop work in his doctoral thesis (2008a, entitled *Unreflective Action*), Rietveld has argued for a picture of human action in which 'we act adequately, yet unreflectively' (Rietveld 2008b, 973). In work with Pim Klaassen and Julien Topal (2010), Rietveld has further extended this argument to social interaction. Here, we will be focusing on his sole-authored work, and return to the application of his ideas to interaction in the next Part of the thesis (in Chapter 19).

The paper focused on here is entitled 'Situated Normativity: The Normative Aspect of Embodied Cognition in Unreflective Action' (2008b). It is the most pertinent for present purposes, for it is here that Rietveld presents the fullest overall account of his view of acting adequately, yet unreflectively. Rietveld's account is important, for it is one of the most thorough attempts to take seriously embodied, affective and dynamic research paradigms in the philosophy of mind and cognitive science, and to use those paradigms to account for the normative dimension of action. The key issue we will be looking at in his work is how he conceives of the normative. As we shall see, Rietveld's conception of the normative has implications for how he pictures the mind learning and working. This picture, which might otherwise gain a great deal from the application of the above important research paradigms, suffers precisely because of the way Rietveld incorporates the normative into his account.

A. Rietveld on the Normative

When Rietveld defines 'normative' in a general way, he says it refers to 'when we distinguish between better from worse, correct from incorrect, optimal from suboptimal, or adequate from inadequate in the context of a specific situation' (Rietveld 2008b, 974). This is a promising start, for it does not immediately narrow the normative down to judgements of correctness and incorrectness, but makes room for other concepts, e.g. better / worse, adequacy / inadequacy, which indicate or at least promise greater flexibility, and thus also hold out hope of describing normative experiences that are not dependent on the prior internalisation of determinants of appropriateness and inappropriateness. Much, however, depends on what Rietveld

means by 'a specific situation', including principally whether he imagines this situation as already organised normatively (in a similar way to Dreyfus). That he tends to think of situations in just that way is exemplified in his assertions that when he refers to persons acting appropriately (for instance, maintaining an appropriate distance in a lift: Rietveld 2008b, 973) he means 'appropriate from the point of view of socio-cultural practice' (Rietveld 2008b, 973), or when he says that what he is interested is 'understanding…the links between unreflective actions of an individual and socially constituted norms' (Rietveld 2008b, 973). The reference to practice and socially constituted norms indicates that Rietveld positions his description of the actions of persons in the contexts of activities (or socio-cultural practices) where what is appropriate is determined by certain rules or norms. Indeed, it is telling that although he begins his paper with examples that may be read more flexibly – such as turning the pages of a book, or helping a fellow pedestrian – for the most of the paper he relies on the context of expertise, or 'specialised skill domains', as he sometimes calls them (Rietveld 2008b, 974). The choice of expert domains is not in itself a problem, for expertise can (as we shall show in a moment) be explained in more or less flexible ways, but it does – at least in Rietveld's treatment (perhaps partly influenced by Dreyfus in this respect) – lend itself to a view where persons (the experts, e.g. expert tailors or expert architects) act in a context where what is appropriate or not is well-established. Expert domains – especially on the back of contrasts between experts and novices, which invite us to think that there is a specific, rule/norm/standard-defined target for the novice to reach, and where expertise is nothing more than a matter of mastering the rules governing that domain – can be all too easily used as contexts where the mind is pictured as overly passive, merely a matter of becoming less reflective (and thus more spontaneous in an embodied / affective manner) in being able to act in accordance with the norms and rules. As noted above, we do not have think of expertise in this manner – we can, for instance, think of experts as having their own individual style, which may not be reducible to acting in accordance with what other members of the profession otherwise treat as unquestionable or unchangeable rules / norms. But expertise can all too easily lend itself to a view where what is appropriate is fixed and settled.

As with Dreyfus, however, so with Rietveld matters are not so straightforward, for in fact there are aspects of Rietveld's view that assist us to make room for a broader, more flexible view of the normative (and thus also a more flexible

view of the mind learning and working). This occurs when Rietveld speaks of 'first-person experience of normative tension', and generally of 'appreciating' something being appropriate or not (Rietveld 2008b, 974). As we shall see in a moment, Rietveld describes the first-person experience in very instructive terms, but when he explains its normative element he nevertheless often returns to referring to 'the normative adequacy of an object' being 'founded…in a socio-cultural practice' (Rietveld 2008b, 974), and thus in speaking of 'how rules compel the skilled individual' (Rietveld 2008b, 974). What we will try to do, in analysing Rietveld's view, is to disentangle his description of 'first-person experience of normative tension' from its subjection to the norms constituting what is correct or not in some socio-cultural practice.

The target, then, of Rietveld's explanation are those moments in which experts – tailors or architects – act adequately, yet unreflectively. He understands, of course, that in any one day, a tailor or architect may engage in many different processes, some of them reflective and some not (indeed, his point is to make room for the unreflective, but not at all at the cost of excluding the reflective). The specific actions he has in mind are 'episodes where the activities of a skilful individual unfold without reflection on his or her part' (Rietveld 2008b, 975). These are episodes where the action 'is not guided by explicit reasoning, nor completely causally determined by the environment, but has its own peculiar type of agency or cognition' (Rietveld 2008b, 975). Nevertheless, the ability in question (which Rietveld sometimes refers to as 'embodied intelligence') comes from somewhere: for Rietveld, it is 'acquired through a history of training and experience in a socio-cultural practice' (Rietveld 2008b, 976). We have seen above how vital it is how one describes the manner in which the mind learns, i.e. whether one thinks of it as simply picking up inclinations / dispositions / abilities to do things in certain ways, or whether one allows for the mind learning something much more inherently flexible / adaptable. Rietveld refers, as noted above, to 'a history of training and experience in a socio-cultural practice', the effect of which, he says, is to modify the relationship between body and world in such a way as to make the individual 'attuned to a familiar environment', one in which she 'is able to perceive and respond unreflectively' (Rietveld 2008b, 976). Later in the paper he returns to the issues of training / learning, and makes it clear that he has a very closed view of what we learn, e.g., he says that 'Our past experience *determines* which possibilities for action attract us. Thanks to this process the craftsman perceives a relevant affordance and is directly motivated to act' (Rietveld 2008b, 991; emphasis

added); and once again later: 'the process of being responsive to affordances is inseparable from the craftsman's concerns' and 'these concerns have been shaped through past learning in his practice and *determine* what shows up as relevant for him in this specific situation' (Rietveld 2008b, 992; emphasis added). The only way in which he tries to make the determinism more plausible is by making room for both implicit and explicit instruction / learning (Rietveld 2008b, 989), but implicit learning is still the kind of learning where one learns to recognise something specific as appropriate or inappropriate (it is just that one is not aware of what one is learning). In short, we learn in such a way, according to Rietveld, that makes us responsive to certain things, i.e. makes it possible for us to respond / react / recognise those things, sometimes by being aware of what we are learning and sometimes not. We will see, in a moment, why this account of learning goes hand-in-hand with a narrow view of the normative (and we will also offer an alternative view of learning, based precisely on a broader, more flexible view of the normative).

One issue that is flagged by the above account of learning is what it is exactly that Rietveld thinks persons are responding to (and thus, on Rietveld's view, also what persons learn to respond to). We saw, above, in the debate between Dreyfus and McDowell that it mattered (at least Dreyfus thought so), for instance, whether one pictured the individual as learning so as to becoming responsive to affordances or to solicitations of affordances (where the latter can, from our perspective, be further understood to be invitations or opportunities issued by the environment when one is relating with it). The way Rietveld sets things up is that he thinks of the affordance as 'an organism's possibilities for action in some situation' (Rietveld 2008b, 976), and the solicitations as the mere phenomenological layer on top – the emotional solicitations to act (what he calls affective allure) are as it were just the flotsam on top of the affordances that are features of the situation. This is important to note because we see something more tentative in the action we are describing when we picture the persons as responding to solicitations rather than affordances (and not as acting within a situation constituted by affordances that are experienced as solicitations). When we think of solicitations instead of (rather than on top of) affordances, we begin to break down the door of thinking of a situation as pre-structured (as when we think of it as pre-structured by affordances because, say, we think of those affordances as constitutive of expertise in some domain). In other words, thinking of solicitations as stand-alone objects of responsiveness helps us avoid any split between the

metaphysical and the phenomenological, and allows us to describe the agent in more active ways. The problem with splitting things into the metaphysical and the phenomenological (as Rietveld does, for he relegates, like McDowell did, the affective allure of solicitations to the phenomenological level: see Rietveld 2008b, 976-7), is that we do not make any room for the person's capacity to be surprised by the invitations / opportunities offered by the environment, thereby creating what appears alluring in the course of action (instead, we picture the individual as always passively merely responding to what we posit – because of how we think of the nature of some domain of expertise – really is an affordance, and is merely experienced as an allure by the individual). Of course, this does not mean that we should see the individual as not being capable of responding / recognising certain things (though even here, as noted in Chapter 11, we need to see this more creatively, i.e. as extending likenesses), but it does mean that we cannot limit our description of action to passive recognition / response. To return to the question of learning for a moment, some of the difficulties that arise with picturing persons as too passive are the result of how one conceives of persons as learning, i.e. if one makes room, at the level of description of learning, of learning not merely to respond / recognise, but learning something inherently more flexible / indeterminate etc, then one may also come to describe any particular act of exercising 'what' one learns in a more open / flexible way.[131]

It may be – to also remind ourselves of another matter that came up in Section 11.C. – that part of the motivation for making a split between the phenomenological and the metaphysical is to 'solve' the frame problem, i.e. show how what is relevant is determined (e.g. in the situation). Rietveld says at one point that 'the skilful individual's responsiveness to *relevant* affordances forms the core of the normative aspect of unreflective action' (Rietveld 2008b, 977; original emphasis), and this indicates a metaphysical reading, i.e. a reading in which expertise in some domain is understood to be constituted by certain affordances, and thus where the person is, as a result, placed in the context of a situation where what is correct (or incorrect if he fails) for that person to do is to recognise those affordances assumed to be established as relevant (which, on Rietveld's view, he learns to do by experiencing their affective allure). The point made, in reply, to this tendency in Section 11.C. also applies here:

---

[131] Language lets us down here, for we immediately speak of 'what' we learn as if there was precisely one determinate thing.

we do not need to 'solve' the frame problem, for the frame problem is instead better understood as a reminder of the very limits of the ambition to fix what is relevant. In other words, we can defeat this way of positioning the individual as passively recognising what he ought to recognise when we refuse the split into the metaphysical and phenomenological, and focus on the agent's experiencing in the course of action, some part of which may involve responding / recognising (though still in a creative way) to what the person has previously treated as relevant, but some of which may also involving creating relevance and solicitation in the course of action.[132]

B. Towards Appreciation

Despite, however, the above drawbacks, when Rietveld does come to analyse an instance of the exercise of expertise, his descriptions are (from the perspective of this thesis) instructive and helpful, especially in terms of getting to grips with what we have been referred to as 'the context of appreciation'. Rietveld's description is largely based on his reading of several of Wittgenstein's passages on craftsmen at work. It is significant that these passages occur in the context of Wittgenstein's lectures on aesthetics (Wittgenstein 1967), for it is arguably precisely in that context that we are more likely to prioritise the context of appreciation (rather than justification) and thus see the mind learning and working in a different way. The key distinction introduced by Rietveld in this context is between 'directed discontent' and 'directed discomfort'. The example (used by Wittgenstein and also relied on by Rietveld) is that of an architect at work, correcting the design of a door. Directed discomfort in this context (and here Rietveld is interpreting Wittgenstein quite liberally, as he acknowledges) occurs when the architect thinks that the door 'just does not seem right to me' and, for instance, destroys the object and starts again, or goes home leaving it behind (Rietveld 2008b, 980). It is where the architect 'is certain that as it is the door is not right…[this is] an experience of overall incorrectness…[it has] an all-or-nothing character; it is undifferentiated' (Rietveld 2008b, 981). Directed discontent, by contrast, is, as it were, more fine-grained or subtler. Here is how Rietveld describes it:

> Directed discontent is related to the craftsman's ability to make all kinds of subtle discriminations instinctively and immediately in unreflective action.

---

[132] Again, this is not an arbitrary 'my say-so trumps' creation, but rather creation in the course of a relation with the environment.

> This enables him to see what should be done to improve the current situation or solve the problem; to perceive and act on possibilities for action (affordances). An example of directed discontent could be a situation in which the architect notices that the door in its architectural context is incorrect and immediately senses two relevant alternatives (make the door more narrow or make it higher), and, what is more, he responds immediately to the best of these possibilities for action (by saying: 'Make it higher', or by an equivalent non-verbal reaction). Here directed discontent seems to be the connotation of having a sense of the right proportions (or other specific aspects) of this object in its context. (Rietveld 2008b, 980)

It is interesting that we are encountering, as we did with McDowell and Dreyfus, some of the most important difficulties for this thesis in the context of subtle discriminations.[133] It is noteworthy that in this passage Rietveld speaks of 'improving the current situation or solving the problem', for we could build on this and try to tease out a notion of improving where the architect is not depicted as merely recognising the way in which it is assumed (as a result of assuming what the expert domain requires) it ought to be improved. As it is, the passage does not do this, but instead speaks of the architect recognising the 'right proportions' or 'responding immediately to the best of the available possibilities', which raises the obvious question: 'right' or 'best' from whose perspective? The point is that the architect is being depicted as passive with regard to what can be right or best; the effect of the way the normative is being conceived is that it pictures the architect at work being driven by his internalised knowledge (learnt by training in a socio-cultural practice) as to what is right or best. There is no room made here (as there might be with the more flexible normative concept of 'improvement') for the architect to introduce / create the way he might come to appreciate 'rightness' or 'best-ness'.

Notice how this also means that the description of the architect's mind is restricted to two possibilities: either too high or too low, or to immediately and instinctively deciding, as Rietveld says later, 'that the door should be 14.8 ft high' (Rietveld 2008b, 981). When Rietveld says 'should' here, he is attaching the 'should' to norms that determine how the door should be, with the result that the architect is pictured passively as merely being able to immediately and instinctively recognise that 'should' (rather than being the author or creator of it, as where the architect feels, from his perspective, that the door should be like this).

---

[133] Recall, from the preceding Chapter, the discussion as to whether 'that shade' had to be a concept, and also whether an 'affordance' was better described as a 'solicitation'.

Recall that it has been previously argued in this thesis that focusing on the context of appreciation might allow us to picture the mind learning and working in a different (and more active, flexible and creative) way than a focus on the context of justification (although to say this is not to give up on seeing the activity and creativity of the agent within that context also). In Rietveld, however, the context of appreciation is read very narrowly. Consider the following:

> Appreciation has a normative character: it normally concerns the correctness of an object for the appreciator within its context. Importantly, because contexts are in general complex it is not possible to make fully explicit what appreciation consists of [citing Wittgenstein 1967, 7]. Appreciation in a situation takes place within a practice and within a culture [citing Wittgenstein 1967, 11]. Appreciation of the object (in context) by the craftsman is normative without being (explicitly or implicitly) guided by rules. Although the tailor may have incorporated the rules of his practice years ago, he normally does not refer to these rules in the situation. (Rietveld 2008b, 979)

Notice, first of all, that what is to be appreciated here is pre-determined; the only concession made is that it may not be fully explicit. Secondly, the key point being made in that passage about appreciation is that it is unreflective; in that sense, it differs from the context of justification (as it is often presented) only by not being reasoned through *ex ante*. Further, appreciation is depicted as the culmination of a process of training and resulting in the internalisation or incorporation of rules. It is not so much a theory of rule-following, for it explicitly states that appreciating may not need to be a matter of being guided either implicitly or explicitly (it is not explained what implicit rule-following may be); it reads more like a theory of rule-accordance or rule-conformity, i.e. to appreciate is simply to act in accordance with – in an immediate and non-inferential manner, and thus in an unreflective or un-reasoned affective and embodied manner – the norms or rules that determine what is appropriate. Indeed, when discussing justification a few pages later, Rietveld says that whereas 'Justification is a reason-giving type of normativity [that] occurs after the fact', 'situated or lived normativity manifest[s itself] in expressions / reactions of directed discontent within unreflective action' (Rietveld 2008b, 981). Here, what allows the label of the normative to attach to appreciation is that it is action in accordance with the norms or rules that constitute the relevant activity or domain (and that are the result of the person having incorporated the rules such that he no longer needs to access them reflectively). The incorporation – or, in other words, the

development of the ability to appreciate – comes from socialisation: 'the craftsman was socialised during a process of training and experience in a community of practitioners' (Rietveld 2008b, 979, fn. 8).

Rietveld here misses the opportunity set up by his engagement with Wittgenstein's example, and by his own reference to appreciation. By presenting the architect as passively related to what is appropriate – never the author / creator of it, but only ever the recogniser (as a result of being trained / socialised) – and by understanding appreciation to be basically a matter of instinctive / immediate (affectively and bodily mediated) rule / norm conformity, Rietveld extends, digs us deeper into, the governance view, rather than helping us out of it. We shall ignore here the problems with assuming that mere rule conformity is possible.[134] Rather, we shall try to focus on the above example and bring out another way of taking the context of appreciation seriously – and in doing so, we shall try to disentangle the otherwise promising reference to the role of affect and the body.

Return, then, to the example of the architect dealing with the door. Return, too, to Rietveld's interesting phrase, 'first-person experience of normative tension' and also to his presentation of 'directed discontent' (an important part of which was the making of subtle perceptual discriminations). The first thing we may notice here is that the making of subtle perceptual discriminations (if one takes 'making' seriously) suggests that there are many possibilities, and that they are not laid out in front of the architect (waiting for him to be recognised) but rather are created by him. This already means we will not describe this case as one where there are two possibilities – making the door narrower or higher (for we can surely be subtler than that) – but many possibilities (e.g. introducing a window into the door, or changing the frame, or making part of the door a curtain, etc – the point is that we, the theorists describing the example, cannot list all the possibilities; there is no fixed catalogue of options before the architect). Of course, any of these possibilities might not be recognised / recognisable as appropriate if by appropriate one means in accordance with the established rules and norms of the socio-cultural practice; this, however, should not make us think that the architect's action is deprived of a normative dimension; the architect is concerned about, cares about, the way the door will appear and work – the

---

[134] We have argued above, in Section 5.B., that that idea is incoherent, for there is no such thing as mere rule conformity – the idea depends on assuming, in an illegitimate way, that the rules are automatically applicable and apply themselves.

quality of the door, you might say, and perhaps the overall quality of the building as well – and this broader, more flexible approach to what can emerge as normatively relevant also helps us to see that he can create possibilities that he may experience as alluring (others may not experience it that way – and this is what makes it risky, the point being that it is precisely the capability to take such risks that is part of expertise).

The second thing that we see is brought out by Rietveld's use of the term 'tension', and here is where his tendency to conflate unreflective and immediate is misleading. Unreflective need not be immediate; it can be the site of tension and uncertainty, or hesitation or exploration. Rietveld presents the case of the architect as knowing immediately / instinctively what ought to be done (where the 'ought', once again, is determined by the norms of the practice). But his own use of the term 'tension' suggests other possibilities: the architect, surely, can unreflectively (still instinctively, but no longer immediately) explore the relative alluring-ness (for want of a better term) or the relative solicitation of the possibilities (which he creates) as he goes along. We present the architect as incredibly passive and un-creative if we describe him as being chained to finding alluring that which he has been socialised into finding alluring.

This last sentence points to the third thing we can notice about the case: the importance of how we describe how the architect learns, and develops the ability to appreciate (in this case, appreciate whether the door is appropriate). As soon as we stretch out the process of appreciation, and thus make room for tension, uncertainty, hesitation and exploration, we also make room for a notion of learning that is not about acquiring the ability to recognise this-and-this or such-and-such,[135] but something more like the patience needed to slow oneself down in order not to reply immediately; the ability to open a space for exploration; or the ability to suspend commitment that something must be so-and-so in order to consider alternatives (all of these can be thought of as ways of initiating, and renewing, a relation with the environment). What is crucial to see is that these processes need not be (though they

---

[135] Recall, by way of contrast, Rietveld's account of learning: 'Our past experience *determines* which possibilities for action attract us' (Rietveld 2008b, 991).

317

can be) reflective, i.e. in order to be unreflective we do not have to picture the person as responding immediately, spontaneously, automatically or mechanically.[136]

What is important to resist is explaining what the architect does in time T, by positing that the ability to do that (exactly that) was learnt by the architect in time T-minus-1: to do this would be to present the mind as merely passive, as but a recogniser. Rather, when we open and stretch the process of appreciation in the way advocated above, we also notice that we need to open the learning process: we see that when we learn, we must be able to learn in ways that leave open 'what' it is that we may have learnt – in ways, then, that allow us to draw on our experiences in order to find new ways of coping / new ways of solving difficulties (this also requires a certain openness as to what one may experience as a difficulty).

There is more to say here, and we will be building on it in Parts V and VI of this thesis. For the moment, it is important to provide a sense of what is being missed when we understand the normative in a certain way. What we see, in Rietveld's paper, is that even when one tries to give voice to the normative in a way that draws on the role of the body and the affects, one's conception of the normative can make one position the mind in such a way that our picture of how it learns and works is very confined. Of course, in offering an alternative reading of the example Rietveld uses (partly by reference to his own terms, e.g. tension and appreciation), we are not suggesting that persons (including expert tailors, architects, etc) must always be the authors / creators of possibilities that they can experience as appropriate or not. The point here is not to shift the entire burden onto the context of appreciation as opposed to the context of justification, or onto the individual as opposed to the social origins of normativity. Nothing that we have said takes away from the contingent usefulness of external constraints – we can use (actively and on occasion) these normative resources in guiding ourselves and evaluating ourselves and others (other tailors, other architects). But these normative resources are still ours and not an underlying order, the pre-existence of which is secured by metaphysical or practical necessity (we see this as soon we remind ourselves – as per the frame problem – that the constraints are not automatically relevant and do not apply themselves). As soon as we see that any such constraints are contingently useful resources, we also see that we have the opportunity, every time we act, to challenge them, and / or to create new

---

[136] Reflection, or deliberation, ought not to be the target: in fact, reflecting can be an incredibly flexible and tentative process.

possibilities that we (and others after us) may (but also may not, hence the risk) experience as alluring. Here, the mind emerges as balancing the reactive and active, as both capable of recognising (though still in a creative sense) as well as exploring in a tentative, non-committal way (though still, crucially, in a way that can be normative, e.g. as when we appreciate something as perhaps appropriate in this way, or that way).

We mentioned above that we would challenge the kind of picture of expertise that Rietveld was offering. What we need to say here is this: expertise may – and this may be counter-intuitive at first – be the kind of level of involvement / engagement / concern / care where one has greater opportunity to create possibilities that others may experience as alluring – and not, as Rietveld presents it (and Dreyfus too) – as the kind of involvement where one is simply faster and more efficient at acting in accordance with what is allegedly fixed or well-established as appropriate. In other words, a large part of developing expertise may precisely be the opposite of becoming more efficient – a large part of it may be about being able to slow oneself down more, to open wider spaces for considering alternatives, to suspend one's commitment to certain ways of seeing for longer, and so on, and thereby to change, rather than reinforce, the discipline / skilled domain / activity, and thus to further refine or re-imagine what is relevant, or what we ought to find alluring / attractive / worthy of concern. Expertise, then, is not exclusively or primarily about acting adequately, yet unreflectively in the sense that the expert does not deliberate 'because he experiences it as good "like this" and has no tendency to act differently' (as Rietveld says of a tailor who is cutting a suit – Rietveld 2008b, 982). Quite the contrary, expertise may be precisely in large part about the ability to create different ways to act, and also being able / being willing to act differently oneself. In other words, it may be in large part about difference rather than sameness, about contingent change rather than the reinforcement of a pre-established order (though without ignoring the benefits to be gained from treating, in the course of pursuing some end, something as an unchangeable means – of disciplining oneself by treating something as the same or as ordered). Put differently, expertise is in large part about the quality of one's relations with the environment – the architect's relations with concrete and steel (or other materials she might use), the tailor's relations with the fabrics, the sculptor's relation with clay, the glassblower's relation with glass, and so on.

What is said here about expertise should make us think differently of certain cases we often rely on when describing expertise. Take cases that bring to mind experts teaching novices, or simply novices learning: these cases are often presented (see, e.g., Meredith Williams's reading of 'the learning situation': e.g. Williams 2009, 19-22) as straightforward examples of the teacher disciplining the student into a mould, and the novice struggling to discipline her mind into that mould, where the novice is seeking to acquire exactly the same body of knowledge that the teacher has 'mastered' (this talk of 'mastery' is also present in Williams 2009, 19-22). Certainly, some of what can go in teaching / learning can be like this, but it is vital that we see that it is possible: 1) that the teacher can learn from the student; 2) and that this can have an effect on the 'mould' (i.e. what the teacher thinks ought to be taught).

It was mentioned earlier that there are parts of Rietveld's discussion that can help us, and that we can build on, in creating an alternative to the governance view. Earlier, it was said that the notion of 'improvement' may be a helpful notion (because it is intuitively more flexible than correctness), as well as the idea that when we consider persons working – e.g. experts – we can focus not on them doing what is appropriate (as this is determined by some determinants), but rather on them 'caring' (as Rietveld does occasionally say, e.g. Rietveld 2008b, 982) about doing a good job, or being concerned with improving whatever it is they are working on (a building, or a suit, etc). Rietveld has some very instructive suggestions here about observing the expert at work (when we understand the expert as caring about doing a good job, being concerned with improving), e.g. looking carefully at the 'expert's shouting in frustration…perhaps even his or her sweating' (Rietveld 2008b, 986). These are excellent suggestions, for they point precisely to the process being a mixture of the familiar and the unfamiliar, the smooth and the rough, the immediate and the prolonged, the open and closed space, harmony and confusion, confidence and hesitation (we will see how this becomes important later, for instance in understanding interaction as requiring both being in and losing balance, and not being in perfect balance all the time).

If one sees the importance of this presence of both balance and imbalance, then one will also see that it is misleading to talk of the importance of the emotions and the body (as Rietveld does) by subjecting them to modes for acting in accordance with and feeling the compellingness of norms. That Rietveld so understands the importance of the emotions and the body is clear throughout the paper – he says, for

instance: 'How is the acting individual compelled by norms? I suggest that besides skills, embodied concerns play a key role' where those 'concerns underlie emotions and affective behaviour' (Rietveld 2008b, 987). Rietveld could say that he is not saying the body and the emotions play no other role – that he is speaking only in the context of how we come to be guided, increasingly more efficiently, by norms when our guidance is mediated by the body and the emotions; but the problem with this reply is that it suggests that as long as we are talking about the normative, the only way in which to speak of the role of the emotions and the body is to think of how they make rule / norm following / conformity more efficient (presuming, of course, that reflection and deliberation is less efficient). As soon as we broaden out the normative – partly by considering frustration and sweat (on the back of a focus on caring and being concerned) – we will see a much more varied role for the body and the emotions.

This critical reading of Rietveld (though also one that tries to disentangle what we have argued is misleading from that which can assist us) should help us see that it is not helpful to cut the cake of problems about the mind into reflective vs. unreflective, or the immediate vs. immediate (especially when one associates reflective with not immediately, and unreflective with immediate). Instead, one of the key issues – this being also how Rietveld situates his inquiry – concerns how one conceives of the normative, for this has deep implications for how one pictures the mind as learning and working. If one pictures the normative in an inflexible, rigid way, as when one thinks that the normative is identified by justifiable claims or assertions of appropriateness and inappropriateness, where what makes those judgements justifiable is their conformity to determinants of appropriateness and inappropriateness, then one is also more likely to picture the process of the mind learning and working in an inflexible and rigid way, trying only to make one's determinism more plausible by speaking of implicit learning, or unreflective appreciation, and so on. As soon as, however, one opens up the normative as an issue of the quality of relations (e.g. the quality of relations between an architect and his materials, a tailor with her fabrics), one comes to see the need for more tentative attitudes – e.g. of care and concern – which, precisely because they can be surprised, can learn to care and be concerned about matters they would not have anticipated it would be possible to care or be concerned about. How one understands 'learning' in this context is important – it must not be a form of learning that determines what we

see (recognise) in the future, but rather a kind of learning that allows us to draw on our past in creative ways (where, in a way, we are enabled to relate with our past). In both cases – whether it comes to learning or exercising certain abilities or attitudes – what we need is more room for frustration and sweating, hesitation and uncertainty. We need to resist the urge of the ideal (including the model of expertise) being easy, swift, fluid, immediate, all the while being guaranteed – offering us reassurance and justification – of being accepted by others as appropriate.

As noted above, later, in Chapter 19 of Part IV of this thesis, we will be returning to Rietveld's ideas, and seeing how, together with some fellow researchers, he has applied his notion of acting adequately, yet unreflectively to social life. For now, we leave this discussion and turn to the final Chapter in this Part, which focuses on tacit knowledge.

Chapter 14. Tacit Knowledge

In the final Chapter of this Part we need to look at a category that has received much attention in the last fifty years: the idea of tacit knowledge.[137] The idea is not a simple one, for there are many different ways in which 'tacit' can be used in connection with the analysis of knowledge. But there is one general strategy or feature common to all these ways, which is indicated already by the very term 'tacit'. The term suggests something hidden, i.e. something that is there, even though we may not realise it and have any way of articulating it (either now – and thus potentially being able to sometime later – or never being able to). Before we go on, we need to say something in general about this strategy, and why it is one resisted by this thesis.

Part of what we have been trying to do in this thesis is avoid putting ourselves in the situation where we need to give an answer to the question at the heart of the tacit strategy – i.e. what is hidden? – for this question tends to make us picture the mind in an overly passive way, i.e. a way in which it is but a mode of or vehicle for acquiring or coming to be influenced by what is hidden (whether rules / norms or causes / mechanisms) and then simply re-enacting them. In that respect, the reference to the tacit might be thought of as a certain kind of explanatory strategy: first something is identified as a re-occurring pattern (a 'kind of behaviour' that has been observed on numerous occasions); second, this 'something' becomes the target of explanation, or in other words, this 'something' is identified as something that must be explicable, where explicable is thought to mean identifying the existence of certain conditions prior to it happening, and thus enabling it to happen (or enabling it to re-occur on each occasion); third, there is a quarrel as to what these conditions are like, and how (if at all) they are experienced, e.g. are they rules or norms,[138] or are they causes or mechanisms?[139]

As the terms themselves suggest, the tacit / explicit debate tends to happen at the third step, but this also means that it is forgotten how crucial the first two steps

---

[137] This is sometimes referred to as implicit knowledge, and it also overlaps to some extent with the idea of know-how, as allegedly distinct from know-that – but we shall stick here with the term 'tacit'. Implicit rule-following was addressed above, in Section 10.F.

[138] If they are, are they explicit in form, or are they somehow tacit, and if they are explicit, might they have been internalised such that they are being tacitly followed – and even if they are explicit in form, might there be other rules or norms that are enabling, implicitly, the explicit ones to be followed, and so on?

[139] If they are causes or mechanisms, are they capable of being made explicit into the form of rules or norms, and thus being meaningfully experienced, or can they only ever be scientifically explained and thus not said to be followable by human beings (but only, say, by computers)?

are. It is the first two steps that set up the problem debated in the third step. The reason why this is problematic from the perspective of this thesis – and especially this Part – is that it positions the mind as very passive, and doomed to first acquiring either the rules or norms, or becoming influenced by the causes or mechanisms, and second to simply re-enacting them. This has the effect that we cannot see how the mind can learn something that is much more inherently flexible (i.e. both in terms of something specific it can use to actively construct similarities, but also something indeterminate that is like an invitation or opportunity to be drawn on and construed in unpredictable ways in the future) and also how the mind can create what may be relevant or important, bringing into view new normative resources. It turns the mind, in other words, into a rather slavish follower of whatever it is that is posited as hidden, and always and necessarily (even without the mind realising it) guiding it or moving it to do something (precisely the 'something' that was identified in the first two steps as the target of explanation requiring a certain kind of explanation).

The above covers a lot of ground in one paragraph, so let us proceed more slowly. In what follows, we shall focus on the most recent, and also arguably the most thorough exposition of tacit knowledge: Harry Collins's *Tacit and Explicit Knowledge* (2010). Although we will focus below on Collins, some of what Collins has to say has similarities with certain views that we will meet later in Part IV (e.g. the work of Bourdieu is perhaps most obviously similar in some ways, and this is discussed in Chapter 19). In that respect, Collins also provides a neat bridge between this Part of the thesis and the next.

There are three parts to Collins's book: the first deals with explicit knowledge; the second with tacit knowledge; and the third with some applications and general discussion of the scheme introduced in the first two parts. Given that Collins very reasonably asserts that one cannot understand tacit knowledge without first understanding explicitness (his way of putting it is that 'The idea of the tacit is parasitical on the idea of the explicit': Collins 2010, 7), we will be examining the first two parts together (the third part shall not be examined here, as the first two parts offer the basic theory). We begin our discussion with what Collins calls 'strings' and their 'transformations', and then go on to look at the three kinds of tacit knowledge he identifies (relational, somatic and collective).

A. Strings and their Transformations

At the foundation of Collins's account is a distinction between what he calls 'strings' (that have certain 'elements', or sometimes what he calls 'information') and other things – these 'other things' may be, for instance, carriers of meaning (such as the words of a language) or forms of instruction (such as rules or hints about how to do things – things that ultimately exist as strings) or forms of explanation (such as scientific explanations). Strings are 'bits of stuff inscribed with patterns', and although they are themselves not meaningful, they are 'anything that is neither random nor featureless' (Collins 2010, 16). The patterns with which strings are inscribed (and which, as noted above, Collins sometimes refers to as elements and sometimes as information) include: 'the 0s and 1s of binary code, the letters of the alphabet, features of the patterns of smoke clouds, electrical states in silicon microcircuits…the repeated shapes on patterned wallpaper, the notes made by a songbird…and so forth' (Collins 2010, 16). Strings are physical things that can have physical impact on 'entities' (entities include humans, but also animals and other living things); being patterns, they can be 'inscribed' on entities 'in many different ways'; and they can change or affect an entity (in the sense of enabling it to do something that the entity could not do before) in one of two ways: first, by 'communicating "mechanically", as when a new piece of code is fed into a computer or a human reacts to a sound in a reflex way', or second, by 'being interpreted as meaningful by a human' (Collins 2010, 17).

Strings are, for Collins, so autonomous as to be capable of reproducing – or being 'transformed', as he puts it – without any mediation by persons (any such mediation is referred to as 'translation', and this may result in 'loss of meaning'; Collins 2010, 25-6). As he puts it, 'A string is just a physical thing and physical things…can be changed into one another and back without loss of the information they contain' (Collins 2010, 26). Examples of transformations include those made by 'electrical signals in the brain…movements of the mouth…vibrations in the air…movements of a diaphragm' (Collins 2010, 28), but more intuitively perhaps, transformations are also what computers do. Translation is the province of humans, and this requires interpretation, i.e. the process of attributing meaning to strings. This process of attributing meaning, as we shall see in more detail in a moment, is a socially-determined process, and thus persons are able to find strings meaningful as a result of socialisation (without society, there would still be strings, but the strings

would not be meaningful; this also means that the strings must be shared, prior to being potentially attributed with meaning by a collective).

How are strings attributed with meaning? Here Collins argues that the strings themselves (because they have no 'inherent meaning') do not determine what they mean, but they do 'afford' certain 'interpretations' (Collins 2010, 35). This is particularly clear in the case of analogue strings, such as a diagram of a house, which do not determine the meaning of 'house', but afford that meaning more than, say, a 'hut' (see Collins 2010, 37-8). Images, including for instance photographs, as well as letters of the alphabet, are meaningless in themselves, but as a result of attributions of meaning by a collective, acquire certain affordances for those members of the collective. 'Affordance' here, then, is not so much an argument for the possibility of seeing the diagram of the house differently – e.g. imagine an individual seeing it, for the first time, as a diagram of expensive houses – for it is still thought that the only reason why the diagram has meaning is because something (i.e. the collective) determined it; rather, it is an argument about how the diagram does not, of itself, determine the affordance, but must be given it by the collective (more is said about this power attributed to the collective in Section 14.E. below).

We noted above that there were two ways in which a string can come to affect us (allow us to do something we could not do before): first, by a mechanical process, as when, for instance, I develop a reflex, e.g. I am trained as a soldier to respond to my superior's commands in certain ways, or I am trained as a slave to row a galley;[140] and second, by a process involving interpretation (and thus attributing meaning). This second process can come in a variety of forms: e.g. it can come in the form of developing an ability to answer a question on a quiz night, i.e. I learn something (e.g. what a hologram is – and I learn it in a pub on the back of a beer-mat) and then I come to be able to answer questions about it (e.g. what is a hologram in a pub quiz; Collins 2010, 58-59); or it can come in the form of different kinds of rules, e.g. in terms of learning to ride a bike, a rule that says 'Try to look well ahead of you when you ride', or a rule that says: 'It should take between half an hour and half a day to learn to balance reasonably well' (which Collins calls a 'second-order rule'; Collins 2010, 62). It is not always clear what Collins refers to by 'socialisation': he seems to want this to apply only to the second process (the second way, and thus the

---

[140] Collins says: 'The slaves are trained through the mechanical effect of punishment and reward and, as far as the slaves are concerned, meaning and interpretation is not involved' (Collins 2010, 58).

meaningful rather than mechanical way, in which strings can affect entities, i.e. us), for he says, for example, 'It is through socialisation that we learn to attribute roughly the same meanings to similar strings most of the time' (Collins 2010, 45), but it is not quite clear how this differs from the first process (where a galley slave or a soldier is 'socialised' into associating certain responses with certain commands).

Before we go on to look at how Collins argues that the above helps us make sense of the distinction between explicit and tacit knowledge, let us make some observations. First, what can we say about the idea of strings? It is, in truth, difficult to make sense of. On the one hand, as we saw above, it is supposed to be the realm of ordered phenomena that has certain features ('not random and not featureless'), but on the other hand, it is also that which can be the object of being made meaningful. Thus, for instance, Collins says: 'Life is not complete chaos because the relationship between strings and meanings is stabilised for much of the time through the sharing of taken-for-granted ways of going on in social groups' (Collins 2010, 45, as when, as a result of the collective determining it, something affords a certain interpretation for us, i.e. the members of the collective). This confuses things, for it suggests that the defence against chaos lies in the socialisation of meaning (in meaning being created by the collective), whereas it has already been said about strings that they are not random and not featureless (the question then arises: not random, and not featureless in what way? How do we know?). Second, what is the relationship between the idea of strings and the notion of the mechanical process by which strings affect persons (or other living things)? Could it not be the other way round, namely that there are processes of stimulus-reflex that create what comes to emerge as a string (i.e. as a pattern) – in other words, the string is not transformed, so much as created and indeed maintained (created because maintained) in the mechanical process? Third, what is the relationship between a string and the interpretative process by which it affects persons (as opposed to living things)? For instance, how is that we come to exercise the interpretive process – attribute meaning – to that thing rather than something else (in other words, is the idea of a string not another way to determine, and thus position persons passively vis-à-vis, relevance)?

The very idea of a string suggests an order that pre-exists human beings. On this view, human beings, whether they know it or not, already live in an ordered way (they are governed by strings). When (and only when) they live in collectives, they attribute meaning to those strings – though mostly they are just governed by those

strings in mechanical ways (via stimulus-reflex, or reward-punishment). One thing that is difficult to see in this is how human beings could ever experience life as chaotic and not meaningful: they would not experience this in the case of the mechanical process, for here the slave just responds to commands; and they would not experience it in the interpretive process, for this process is by definition a matter of attributing meaning (to something, in any event, that is already ordered). Another thing that is difficult to see is how persons can come to create order for themselves; it seems that the only kind of 'order' that they ever 'create' is the attribution of meaning in the interpretive process, i.e. a kind of phenomenological layer on top of the already ordered transformation of strings (i.e. on top of precisely the mechanical processes of transformation of strings).[141]

The result of all this is that the mind is pictured as passive in two stages: first, the stage at which it is subject to, or determined by, the transformation of strings (e.g. by a mechanical process of stimulus-reflex), and second, by the stage at which it comes to interpret certain strings in certain ways, for the way in which it does so is determined by the collective (it is only thanks to the collective that there are affordances in the first place). The account is neat, for it imposes a social order onto a 'natural' one: the meaning given by the collective (thereby constituting the social order) is only ever attributed to that which is already a string (already something not random, and not featureless, and thus a kind of natural order). This guarantees that social order is also causal, and thereby avoids problems associated with 'free-floating domains of meaning' (or with the problems that bedevil claims that it is rules or norms all the way down). But the cost, as noted above, of this neatness is that it presents the mind as passive in two ways, where both whatever it does as well as what it experiences as meaningful is always and already determined; persons, in other words, are depicted as always and necessarily governed by either or both the mechanical and the interpretive processes of the transformation – perpetuation – of strings.

This view can be unravelled – and thus the picture of the mind as passive in two stages can be resisted – when we take another look at the notion of strings, or more broadly, patterns. Turn back to the idea of the diagram of a house, or indeed the presence of a photograph. It is true that we can have the experience of seeing either a

---

[141] On Collins's view, to recall, this phenomenological layer is only ever available because of the collective.

series of lines or a house / person (imagine the photograph is a portrait). Collins effectively takes these two experiences and elevates them to a level of a general theory: the patterns always below, and the meanings sometimes on top. But the category of patterns here is only meaningless when it is contrasted with the experience of patterns as meaningful, i.e. the patterns are not meaningless in themselves. In other words, I need not see something as patterned in some way for me to experience as it meaningless. Seeing something as a series of lines can be very meaningful.[142] It being patterned is not what makes it meaningless, for I can also be looking at a diagram of a house or a person in a photograph and I can experience this to be meaningless to me. Collins is right that I can have the impression, when looking at a diagram for too long, say, that what first struck me as a diagram of a house no longer functions as that diagram – the diagram does not speak to me as a sign of something else. But it seems to be going too far to say that any object I take as meaningful is breakable down into one, and only one, pattern – this being the underlying pattern that is meaningless until the collective ascribes meaning to it. There are not meaningless patterned objects on one side and meaningful ones on the other, but just two different experiences: one where we see things as meaningful, and another when we see them as meaningless.

What needs to be avoided here is the elevation of the patterns to a metaphysical category, as if, for instance, there was but one way in which a house was patterned, this way being the 'underlying' meaningless pattern. One can see something as patterned – break down a face, say, into some features – but one can also see it, as it were, holistically, and both the face and some features can be experienced as either meaningful or meaningless (imagine I am comparing features of faces). Perhaps it helps to see that I can also treat a face as a meaningless feature of an object with meaning for me, as when what is meaningful for me is the presence of a crowd, and not any particular face (if, in these circumstances, I focused on a particular face, then given that I am experiencing the crowd as meaningful, that face might strike me as meaningless). The point here is that what we may treat as a set of

---

[142] For a fascinating cultural history of 'lines', full of pictures of lines capable of being experienced as meaningful, see Ingold 2007. The idea here is that the form of the lines is experienced as meaningful – precisely because the comparative pool are pictures of lines. Perhaps what can say is this: something can be meaningful when we treat it as a sign for something else, or when we look at its form (e.g. when comparing its form with some other form), or even more generally: when something cannot be used in some way its meaning is threatened. In any event, the idea is to resist the argument that there is some category of things that is, as it were by nature or essence, meaningless.

meaningless properties together comprising something meaningful does not have metaphysical status as an underlying pattern because: 1) there are many other ways of seeing that which is experienced as meaningful as being patterned (there are other sets of properties one can invoke, other ways in which one can break down the object into different features); and 2) what can sometimes function as a property or set of properties can at other times function as precisely the object of meaning. We are talking about the phenomenology of meaning here, and not metaphysics: there are no underlying, fixed patterns to which we but attach meaning. We are not, as it were, doomed to living in two determinisms: first, the determinism of the natural order (comprising of underlying patterns); and second, the determinism of the social order (comprising of affordances attributed by the collective). Collins only comes to think so because he over-generalises a particular experience (of seeing how something can be broken down into lines experienced as meaningless in light of the meaningful whole), leading him to restrict the (collective) mind's work to attaching meaning to (what he thinks must be) underlying patterns. In short, we need to see that strings themselves are products of, and can be changed by, the mind: there may be an argument that certain ways of ordering (e.g. the alphabet) are more stable than other ways of ordering (e.g. particular individual styles of linguistic expression, or literary genres, etc) but this relative stability ought not to dichotomised into the meaningless foundation of patterns and the collective ascription of meanings (called by Collins the interpretative transformation of strings) onto those patterns.

B. The Four Meanings of 'Explicable'

Let us return now to Collins and see how his notion of strings and their processes of transformation relates to his discussion of explicit and tacit knowledge. Collins's understanding of the explicit is, in fact, incredibly broad. It is perhaps best articulated in the following 'four meanings of "explicable"':

1. Explicable by elaboration: A longer string affords meaning when a short one does not;
2. Explicable by transformation: Physical transformation of strings enhances their causal effect and affordance;
3. Explicable as mechanisation: A string is transformed into mechanical causes and effects that mimic human action; and
4. Explicable as explanation: Mechanical causes and effects are transformed into strings called scientific explanations.

(Collins 2010, 81)

This is very broad, for there are four quite different things here: there is a thesis (in number 2) about how the causal effect and affordance of strings can be enhanced (basically, it seems, by being repeated and past down the generations); there is a thesis about the meaning of strings (in number 1); there is a thesis about how the 'same' results can be achieved by certain mechanisations; and there is finally one widely used notion of explicate, which refers to 'scientific explanations' (in number 4). It is difficult to see how number 2 can be a meaning of 'explicate'. Number 1 is difficult to assess, for it seems to be of very limited scope: it refers to the phenomenon of one string standing in for a longer one, where the longer one affords a certain meaning[143] – but it is unclear why this is an explication as opposed to an abbreviation, of something shorter standing in for something longer.

This leaves us with numbers 3 and 4. It is not clear why number 3 should be counted as a case of explication: it seems to be a matter of creating a mechanism that can achieve what one counts as the 'same' result achievable by a human being – but it is one often referred to in Collins's discussion of tacit knowledge that follows, so we shall postpone discussing it until we discuss the examples where it is said to apply. Number 4, on the surface, appears to be very narrow, for it refers only to scientific explanations; in fact, as we shall see in a moment when we look at some examples, Collins includes in number 4 the kinds of instructions and hints we may make in teaching someone how to ride a bike (i.e. it includes the first-order rules – such as keep your head up, etc – as well as the second-order rules, these being the scientific ones, e.g. about how long it ought to take you to learn to balance, etc). The key thing about this way of thinking about the explicit – which we will examine in more detail in a moment – is that they take there to be an unproblematic explanatory target: in number 3, the target is the 'same result'; in number 4, it is some specific way to do something (e.g. some specific way to ride a bike). These two targets are in fact the same, and what numbers 3 and 4 offer are different ways of 'explicating' them (to use Collins's terminology): number 3 tries to find a mechanical way to get the same result, and number 4 offers a scientific description of the process needed to get the same result or a set of instructions needed to get the same result. The problem, as we

---

[143] The best example of this in Collins's book is where the letter A is said to be a string, but one which is broken down into a longer string via the binary code '0100000010' (Collins 2010, 16).

shall see, will lie in the assumption that there is a 'same result' (i.e. precisely an unproblematically identified, or fixed, end).

Against the background of these notions of 'explicate', Collins offers three categories of tacit knowledge (i.e. knowledge that cannot be explicated): first, what he calls 'relational tacit knowledge'; second, 'somatic tacit knowledge'; and third, 'collective tacit knowledge'. Let us examine each category by reference to the examples offered.

C. Relational Tacit Knowledge

The first category (relational tacit knowledge) refers to the 'contingency' sense of the 'cannot' in 'knowledge that cannot be explicated'. This, then, is allegedly knowledge that although it could be explicated, is not. The leading example is of knowledge that is 'concealed': 'Knowledge…is sometimes kept secret by elite groups. Secret societies have certain rituals that are kept hidden so as to exclude outsiders…the knowledge is vouchsafed to those within the appropriate social network or social space…but hidden from those who do not belong to the in-group' (Collins 2010, 92). This is a very strange category of the 'tacit', for when theorists refer to 'tacit knowledge' they usually speak about the kind of knowledge, not about whether it can be accessed as an insider or an outsider to some group. This strangeness is highlighted when one considers whether the knowledge Collins is thinking of is, say, written down and locked away in a safe, or whether it is another kind of knowledge (but then what kind?) – the point is that this question is what is normally taken to be constitute the issue as to what is the difference between tacit and explicit knowledge. In other words, the notion of concealment is neither tacit nor explicit, for presumably either tacit or explicit knowledge could be concealed. Notice, also, that if concealed knowledge were to be a kind of (relational) tacit knowledge, it is unclear how it would correspond to any meaning identified for explicating: none of the 4 'meanings of explicate' above appear to be applicable to revealing the secrets of one's company or secret / elite society.

Another example of relational tacit knowledge is what Collins calls 'logistically demanding knowledge.' His example is that of a traditional manufacturing company that has 'an warehouseman who knows where everything is', and where the point is that 'the warehouseman could be replaced with a computerised system so long as someone devised a numbered set of racks and a catalogue with

enough classifications of the objects on the shelves to make them easy to recognise and match with other objects' (Collins 2010, 94). About this example, Collins says: 'the choice of which way to go – stick with tacit or go to explicit – is a contingency of time and place so the knowledge, if it is kept tacit, is relational tacit knowledge' (Collins 2010, 95). This, too, it must said, is a strange notion of tacit knowledge, for it presupposes, rather than explains the notion, i.e. it presupposes that the warehouseman's knowledge is tacit while the computerised system would be explicit. This only works if it could be said that the computerised system 'makes explicit' the warehouseman's knowledge, but of course this is precisely the issue: in what sense exactly would the computerised system merely make explicit the warehouseman's knowledge, as opposed to assuming it gets the 'same result' via an alternative means?[144]

Another example of relational tacit knowledge is 'unrecognised knowledge', i.e. knowledge that we do not yet recognise as knowledge, but that we might some day discover was knowledge (as when we discover some 'previously unnoticed parts of the procedure' for, say, reproducing suspension fibres: Collins 2010, 96). This, again, tends to revolve around examples where we want to recreate some effect, and where we have some data about this effect was previously reproduced. The idea is that we study this data, and one day we 'discover' something about the procedure we had not previously noticed (something we had not – and also those that had performed the procedure had not – 'made explicit'). The problem with this example is about what we claim to have discovered: in what sense can we claim that we have 'made explicit some previously unnoticed parts of the procedure', as opposed to simply finding a way to reproduce the effect? Further, on what basis can we be sure that we achieved 'the same effect'? Finally, on what basis can we say that we have finally 'discovered', 'the previously unnoticed part', i.e. how do we know we have identified 'the action', as opposed to part of the action (which we only 'discover' later)? These two last questions are related. The claim (that we have discovered the missing part not previously noticed) depends on us being sure that we have reached the same effect – but imagine, later down the line, that we 'discover' that in fact the suspension fibres we created ought not to be treated as the same, e.g. they crack under a certain amount of pressure, the point being that we could not have known that earlier, for we had not

---

[144] We shall say more about this issue when we turn to somatic tacit knowledge.

used that criteria as criteria for determining whether we created the 'same effect'. As soon as one sees this, one can also see that what one may have 'discovered' need not at all be 'the missing part', for what one may have discovered may have, for instance, been a part of a bigger action – imagine you 'discover', later in time, that whereas what you thought was 'the missing part', e.g. a shake of the test-tube with one's hand, was – in light of the discovery you made about the certain amount of pressure under which the fibre cracked – a shake of the test-tube with one's hand as one walked along at a certain tempo. This is important, for it shows that the claim of 'making explicit' is dependent on keeping certain matters fixed (i.e. keeping fixed what counts as 'the same result' and what makes the 'discovery' a successful identification of 'the missing part').

We have to conclude, then, that relational tacit knowledge is either not at all a category of tacit knowledge (as in the example of concealment), or that it does not show why the mode of the explicit (mechanising the warehouseman's job, or discovering the missing part) is in any sense making explicit something that was there.

D. Somatic Tacit Knowledge

The warehouseman example is a good way into somatic tacit knowledge. The idea here is that there is a kind of knowledge that 'becomes established in our neural pathways and muscles in ways that we cannot speak about' (Collins 2010, 99). This category builds on the famous example offered by Michael Polanyi: the ability to ride a bike.[145] In the case above, the warehouseman may 'know' how to find the objects people ask him for, but he may not be able to explain how he knows. The way that Collins deals with this category is to say that it is knowledge that arises because of our 'bodily limits even though it can be explicated in the third and fourth senses of explanation' (Collins 2010, 101).

One way in which Collins seems to think that the knowledge in question is tacit is that it is a matter of having learnt to do something by way of, but now no longer needing to consciously access, a certain set of rules; in other words, he equates 'tacit' here with something more efficient (i.e. no longer requiring the slow process of deliberately accessing the rules). The problem with this is that it assumes that the

---

[145] 'If I know how to ride a bicycle…this does not mean that I can tell how I manage to keep my balance on a bicycle' (Polanyi quoted at Collins 2010, 101; see Polanyi 1958).

expert is unconsciously following rules, and this is of course very problematic (as we have been discussing throughout this Part).

Another way in which Collins seems to think that the knowledge in question is tacit is because he thinks there really are rules or mechanisms for balancing on a bike, even if we do not know them. He seems to think this is so because it is possible that we will one day be able to build machines that can do many of things we do, e.g. play golf or chess ('nearly all sports…could, in principle, be executed by machines': Collins 2010, 106). The point he wants to make here is that the rules or mechanisms that we may use to make the golf or chess playing machine may not be ones that human beings could consciously access – but us not being able to access them does not mean that our golf playing or chess playing cannot be made explicit (explicit in the third sense, i.e. mechanisable). This is an argument repeated a number of times, but it is very problematic. The basic problem with it is that it assumes the machines would be doing what human beings do when human beings play golf or chess – otherwise, what sense could be made of the claim that the mechanisation is a mechanisation (a form of the making explicit) of human golf or chess playing? But in what sense would the machines be playing golf or chess?

Collins in fact makes two claims here: first, if we think of playing chess (let us put golf aside for a moment) as simply 'the ability to win' (Collins 2010, 108), then we can build machines that are able to win at chess; second, if we think of playing chess in terms of following some specific process (e.g. not calculating variations, as a machine does, but recognising patterns) then we can also build machines, for 'there is nothing…that would prove that the human way of chess playing could not, in principle, be described in terms of a mechanical sequence instantiated in the particular substances of brain and body' (Collins 2010, 110). The same problem arises for both claims: in order to say that the mechanisation makes explicit human chess playing, one needs to define what playing chess involves, whether one does so solely by reference to the 'same result' (winning), or solely by reference to following the 'same process' (e.g. pattern recognition). Collins knows there is a problem with using 'the same result' as the criterion for playing chess, but the problem he identifies – that computers play a different kind of chess than humans – is not the major problem: the major problem is in thinking that there is one result, or as it were one way of identifying the aim, of chess playing. This assumes that the game of chess is somehow constituted by – could not be played if one did not aim for – such-and-such

an aim (winning the game). This is deeply problematic, for it would mean that we would not understand, for instance, how there can be so many draws in tournament chess (sometimes, one is indeed aiming to draw, not to win the game), or how two persons could sit down to play a beautiful game (and not worry about the result), or how two persons could sit down to play a training game where the point was to experiment with a certain set-up in the opening to see if it suited the style of the player. The same problem arises with the idea that we could build a machine that plays chess by 'pattern recognition': for in what sense does pattern recognition capture the process of playing chess? Consider, just as a starting point, the differences in playing styles between leading chess players: some prefer closed positions, and some open positions, and some prefer attacking and some defending. Surely, these stylistic aspects are parts of the process of playing chess, but does the idea of 'pattern recognition' capture them? If it does not capture them, in what sense could it be said to be mechanising (as a form of explicating) human chess playing?[146]

The real problem here, then, is that Collins is fixing (or holding still, or defining in some closed way) what it means to play chess (cut up in terms of some result and some means for achieving it). Chess, as we have seen many times in this thesis, is a dangerous example, for it is often (mis)understood as having a certain nature (where, say, winning is the aim, and where one must follow such-and-such rules) – so it is all too easy to (mis)use as an obvious case of something mechanisable. But consider the warehouseman example: here, it seems more difficult to cut up what the warehouse man does into some definite end and some definite means. Imagine that one says that the end is getting the right object: in a way, that is true, but of course the point is that the warehouseman gets faced with various kinds of demands (e.g. showing some part that has been mangled into an almost unrecognisable shape), where his ability may be one that involves having seen many mangled examples of all the parts so that he can (creatively) recognise this one as close to that (perhaps he brings several possibilities in the first instance, rather than 'the right' one).

The point is that the end and the means are not so obvious: they only become obvious precisely when one mechanises them, i.e. one puts labels on everything, and then asks the machine to get that precise make and model (in way, to assume

---

[146] Perhaps the difference between humans and machines is 'the demarcational problem' for the twenty-first century, in the way that the difference between animals and humans was for previous centuries (for few now hold the view that animals are just machines).

something is mechanisable one must assume that one will not lose anything in the process of mechanising it). The warehouseman does not necessarily follow the same procedure each time – there are not some means laid out for him that he only follows to get to a certain end; he may need to invent a new way of solving some problem (someone comes to him, say, with a part of a part, and that has never happened before; perhaps, sometimes, he tells them: the problem is not with the part, but with the machine – it will keep on spitting out mangled parts). The same goes with a chess player: a chess player does not always solve every problem on the board by pattern recognition; some of the problems, e.g. what first move to play, have nothing to do with pattern recognition at all (choice of opening concerns more the peculiar taste one might have for open or closed positions, or manoeuvring or tactical positions, etc). The point is that a mechanisation, or indeed a scientific explanation, places its bets on a certain understanding of what is going on (what the warehouseman's job consists in, what chess playing involves) and sets up a machine on that basis: but any such 'understanding' is bound to be very narrow (limited in its specification). Put another way: a machine has to be re-programmed when confronted with something that does not conform to the pre-established means given to it for solving certain pre-defined problems; a human being does not, for a human being's ability (to play chess, be a good warehouseman) is not so pre-defined / pre-established.

There are two possible sources for why Collins does not see these possibilities (and thus thinks playing chess or riding a bike is explicable – mainly in the sense of mechanisable): first, because he thinks (in a way that does not seem problematic to him) that there are underlying patterns (or strings), and thus does not realise that he is creating his explanatory target (e.g. he might think there is one way of balancing a bike, whereas how one balances a bike may depend on not only one's body and one's abilities, e.g. muscles etc, but also on an indefinite range of possible circumstances, e.g. imagine confronting, for the first time, riding down a hill); second, because he thinks that certain activities or tasks (or jobs etc) are constituted in certain ways, such that they have a nature, and such that there is only one way of doing them and only one reason for doing them that way.[147] For it is only if one thinks that there is just one way of doing something, and one reason (one aim) for doing it, that one will think it is

---

[147] That this is indeed what Collins thinks about activities / tasks etc is obvious from the third category of tacit knowledge, the collective category, which is discussed below.

mechanisable – as if human beings (especially the non-experts) were just very inefficient vehicles for re-enacting certain means for achieving certain ends.

E. Collective Tacit Knowledge

The third category of tacit knowledge is collective tacit knowledge. Here, the claim is the strongest: this sort of tacit knowledge is not explicable. But it is not explicable in a very specific sense, which we need to explore (exploring it will also point us to some of the claims we will meet in Part IV of this thesis). We have already seen the collective referred to extensively in Collins's understanding of tacit knowledge; it is thus no surprise to see him state that collective tacit knowledge 'is the irreducible heartland of the concept' (Collins 2010, 119). Some of the examples we have met above reappear in this context: e.g. bike riding and chess playing. Here, however, Collins adds the idea that when these activities 'are carried out by humans' ('carried out' is already very passive talk) 'in the normal course of social life, there are additional aspects of the activities' that need to be taken into account (Collins 2010, 120). What is crucial here is how Collins characterises these 'additional aspects'.

Take bike riding: 'normal human life bikes are ridden on roads used by others and traffic has to be negotiated' (Collins 2010, 120). At first, Collins speaks very promisingly here of the need for 'judgement' (e.g. when driving on 'real roads' and negotiating traffic), but this quickly turns into a picture of what he calls 'social judgement', which in the context of normal, everyday bike riding would involve 'understanding social conventions of traffic management and personal interaction', e.g. 'knowing how to make eye contact with drivers at busy junctions in just the way necessary to assure a safe passage and not to invite an unwanted response' (Collins 2010, 121). For Collins, being able to recognise and re-enact social conventions (including keeping track of any changes in them) is the crucial element in 'social judgement'. Whether it is bike riding or chess playing, both of these activities are constituted socially, i.e. there is a way of doing them appropriately that is defined by conventions: 'The right way to do things can only be captured through experience, and that experience and its application vary from country to country. The explication of the way such things are captured through experience is the socialisation problem' (Collins 2010, 122).

It is important to see how far Collins pushes the idea of social conventions, or what he also refers to as 'social rules'. Thus, for instance, he recalls an episode of Star

Trek when Data ('a clever android') is shown (only once) the 'steps of a dance'. Whereas we (humans) would need to acquire somatic tacit knowledge of the dance (here, somatic tacit knowledge appears to mean merely learning the steps of the dance such as to be able to execute them without conscious attention to them), Data was able to imitate the steps immediately. Collins has no quarrel with this so far, but then he says that the episode went wrong when it showed 'Data managing improvisation as flawlessly as he had managed the initial steps', and here is how Collins explains why this is problematic:

> But improvisation is a skill requiring the kind of tacit knowledge that can only be acquired through social embedding in society. Social sensibility is needed to know that one innovative dance step counts as an improvisation while another counts as foolish, dangerous, or ugly, and the difference may be a matter of changing fashions, your dancing partner, and location. There is no reason to believe that Data has this kind of social sensibility. Social sensibility does not come through having a quickly calculating brain, it comes through having the kind of brain that can absorb social rules… Data can master the steps after the fashion of a computer rather than in the clumsy way that humans would have to master them. But this won't work when it comes to the rules of improvisation. (Collins 2010, 123)

This explanation is promising until we reach the second sentence. Here, already we are told what the sensibility consists in, namely in learning to recognise what step dances count as foolish, dangerous, etc, which means learning the 'social rules'. Even improvisation, then, is conceived of by Collins as a matter of being rule-governed.

Collins, it ought to be added, does conceive of change: there can be change, as is noted by him in the above quoted paragraph, in fashions for example, but when he refers to change it is always a change imposed by the collective, such that the question becomes: 'by what mechanism do humans stay in touch with society and how can one build a machine that can do it?' (Collins 2010, 124). Change is never depicted as introduced by individuals, or experienced by individuals (or indeed really experienced by groups): rather, whenever referring to change, the change is assumed to have already happened (it is not explained how things change, e.g. how language changes), such that the focus of the inquiry becomes how to socialise persons so that they can keep up with these changes (learn the new rules; see Collins 2010, 129).

The crucial part of Collins's discussion is that he pictures social life as a collection of distinct activities, each constituted by certain rules (the right way to play

chess, i.e. follow the right means for the right end), and further 'embedded' in rules (or conventions – he uses the two interchangeably) as to how to 'carry them out' in each culture and to some extent context. Laid out in front of human beings is always a right way of doing things, and this right way is determined by certain determinants (rules or conventions). The result is that the mind is pictured as a sponge for absorbing social rules and conventions – which, coincidently, is also how, according to Collins, we differ from animals and trees: 'humans differ from animals, trees, sieves in having a unique capacity to absorb social rules from the surrounding society – rules that change from place to place, circumstance to circumstance, and time to time' (Collins 2010, 124). Here, we get a hint that even sensitivity to context / changing circumstances is a matter of sensitivity to rules – this is confirmed later, as when Collins says that 'there still remains the problem of determining the current context and being sensitive to changing context' (Collins 2010, 130). Collins imagines that there are rules that determine what context is in front of us, such our ability to be sensitive to context is itself a rule-governed ability (but one acquired and exercised tacitly) – which is, in effect, to try to solve the frame problem (rather than see it as something that indicates the limits of a rule-governed account).

Collins's discussion, then, presupposes the existence of social rules: all of social life is already governed by them, and each situation we confront is so governed. Given that we participate in these successfully (i.e. largely conform to the rules) – this, of course, is a problematic assumption as we have seen (it only looks unproblematic when you assume the automatic relevance / automatic applicability of certain rules) – so it must be the case that we learn them and access them (or act in accordance with them) in ways that do not need deliberation or calculation. Here, we come to see how, despite having identified four uses of 'explicate', and argued that this is how he will think of explicit knowledge, what comes to be counted as explicit here is learning to do something (where the 'something' is imagined as constituted by rules that define the right way of doing it) without learning the rules that govern it, and then acting in accordance with the rules without accessing them deliberatively: 'It is only humans who possess what we call "socialness" – the ability to absorb ways of going on from the surrounding society without being able to articulate the rules in detail' (Collins 2010, 125). The rules, then, are always hidden below. Here, at the 'heartland' of the concept of tacit knowledge, we see how it works: human beings,

insofar as they participate in social life, are tacit learners and followers of the rules that necessarily govern social life.

The resulting picture of the mind is not only passive, it celebrates perhaps the apex of passivity: human beings are referred to by Collins as 'parasites'. Collins says: 'The special thing about humans is their ability to feast on the cultural blood of the collectivity in a way that fleas feast on the blood of large animals. We are, in short, parasites, and the one thing about human brains that we can be sure is special is the way they afford parasitism in the matter of socially located knowledge' (Collins 2010, 131). It is not easy to find a picture more diametrically at odds with what this suggested in this thesis: certainly, it is true that sometimes human beings, perhaps especially when they do things together (such as play chess together, or ride bikes in each other's company) will either need to agree or treat, in the course of interacting in that instance, as unchangeable and unquestionable certain ways of doing things (some of which we may think of in terms of conventions, norms or rules), but none of this should make us deprive human beings (including in groups) of the ability to create possibilities (sometimes completely accidentally) that are nowhere envisaged by the conventions, norms or rules, and the ability to change what they think of as 'the right way of doing things', and doing all this actively rather than passively. It would probably be going too far to turn the picture around, and argue that it is the conventions, norms and rules that are parasites, feeding on persons whose attitudes keep them alive (for the image and language of 'parasites' etc is not a pleasant one). In any event, the more important thing here is to see just how reductive a picture of the mind learning and working we get when we understand social life to be so rule-governed so as to mean that every human action is described as a matter of detecting and acting in accordance with the right paths laid out in front of persons.

You might very well ask how Collins answers the question he poses above, namely whether mechanisation of social sensibility (as he defines it) is possible. This is another way of asking in what sense Collins thinks that collective tacit knowledge is tacit, or not explicable. Collins has a problem here because having conceived of social life as rule-governed, he is forced to limit the tacit knowledge in question to forms of acquisition and then forms of access to the rules: this is why he says that what we are talking about is a 'matter of the nature of humans not the nature of knowledge' (Collins 2010, 138). Here, he shows quite clearly that he already assumes what the structure (or, as he puts it elsewhere, the 'ontology of knowledge, Collins

2010, 96) is, namely that it is rule-governed. That is why he argues that the 'deep mystery remains how to make explicable the way that individuals acquire collective tacit knowledge at all' (Collins 2010, 138; this is what he calls 'the socialisation problem'). His final answer is that we do not have machines that could 'mimic' human social action, and – at least at this stage, and 'until we have solved the socialisation problem' – we cannot 'foresee how to make machines that have the ability to be social parasites, nor describe the ability scientifically' (Collins 2010, 138). But notice that Collins leaves open the possibility that as soon as we 'solve the socialisation problem', and improve our ability to mechanise, it is not out of the question that we could invent machines who could mimic social action.

Let us make it very clear that our position here is one where we do not recognise there to be a 'socialisation problem', and least of all one that can be solved. Like the frame problem, what Collins calls the socialisation problem shows the limits of the approach he takes, i.e. a kind of social determinism (at least in this context of collective tacit knowledge – in other contexts, we are presented as determined by strings, or as parasites on natural rather than social patterns). There is no socialisation problem because there is no one right way – that can be broken down into the right aim, and the right means for achieving that aim – of engaging in social life, e.g. there is no right way to ride a bike, no right way to play golf or chess, no right way to maintain distance between oneself another passerby in the street (as Collins assumes there is: see Collins 2010, 130). Collins wants to see these matters as so rigid that they could be mechanised: broken down into set paths for reaching set targets (when they change, they are somehow changed by the collective so that the individual merely has to 'keep up').

In effect, Collins takes what can often be very useful guides, some that we may actively treat as relevant and unchangeable, and he (over)generalises their importance, transforming them into a theory of human nature and all of social life: of persons as parasites with little more than the ability to absorb those rules, and of social life as little more than collections of rule-governed activities and contexts. The reason why this is an (over)generalisation is that activities, such as riding a bike, playing golf or chess, are not automatically governed by such-and-such rules: it is still up to us to take these as relevant and applicable; even when we do, it does not follow that there is not more to playing golf or chess than the rules that we may, on any one

occasion, associate with the game (for more on this reading of games, see Chapter 7 above).

F. The Very Idea of the Tacit

We have seen that Collins has attempted to uncover several kinds of tacit knowledge. What we found was that each category presupposed that the relevant activity or task (whether it be bike riding, chess or reproducing fibres) was rule-governed (there being the right end to which there was a right means), such that what 'tacit' ended up referring to (especially in the second and third categories, somatic and collective tacit knowledge – because the first, relational knowledge, was more about keeping something concealed) was how to describe humans so as to be able to consider whether machines could do the same thing, i.e. how to describe the way in which humans acquire the rules without being aware of them, and how they access them (and act in accordance with them) without realising it. Strangely enough, when understood this way, we lose sight of any distinction between explicit and tacit knowledge: we end up, instead, with a view of the structure of knowledge (i.e. being rule-governed) on the one hand, with a view about how persons acquire and access (and act in accordance with) these rules on the other hand. The point is that there are not two kinds of knowledge presented: just one, and the only issue is how persons learn to (re-)enact it. The result, as we have seen, is to picture humans not only as passive vis-à-vis normative resources, but as parasitical on them.

One observation to make is that what tacit knowledge – especially in Collins's notion of collective tacit knowledge – depends on is commitment to 'sameness': of first constructing an ontology of social life (of the one and only right way of doing things), and then introducing human beings as a second step. The 'ontology' is kept fixed – understood to be the same (as with the transformation – rather than translation – of strings in Collins's scheme) – for the purpose of describing the epistemology. As a result, persons are depicted as passively confronting an already laid out or organised context or situation, such that their minds are described as merely able to recognise or detect the right means for achieving the right end. But the foundations here are shaky, i.e. the idea of things being 'the same', a commitment to positing and keeping what may be relevant pre-determined and fixed, the assumption that we are just presented with change as a finished product, are all – as has been argued at length above – very

doubtable moves. What has been argued specifically in this Part is that such moves have very deleterious effects on our understanding of how the mind learns and works.

Claims concerning tacit knowledge may sometimes appear attractive or even irresistible because of the way cases (examples of it) are set up. These cases tend to be ones where we might already be inclined to recognise something as a normative resource (a convention, norm or rule), e.g. driving on either the left or the right hand side of the road. This is further reinforced by examples of a contrast: in this country they drive on the right-hand side, and in this on the left-hand side. This makes it look obvious, and uncontroversial, that such-and-such a country is governed by such-and-such conventions (going on from there to have reason to keep the matter fixed, and focusing on how persons acquire and access the relevant convention). But we can unravel these cases in various ways. For a start, we can give an example of something that we would not be likely to already recognise as a convention, norm or rule, e.g. imagine listening to the radio (a radio station called 'Nostalgie') and at one point hearing a jingle along the lines of ('Nostalgie, 9 letters, the station for you');[148] hearing this, one might very well wonder what the '9 letters' is doing there; after a while, one may realise that 9 letters refers to the number of letters in the name of the station ('Nostalgie'); but as soon as one realises this, one's experience of strangeness is not likely to leave one, for why is it relevant or important to have a jingle that points out how many letters there are in the name of the station? Now you might try to describe this case as one where this appears strange because there is (in some tacit or implicit form) a convention, norm or rule that we not refer, in radio jingles, to how many letters there are in the name of the station, but the case is pretty weak: in what sense is there a convention, norm or rule implicitly or tacitly there? Notice that I may have not considered the jingle as strange – I may not experience it as inappropriate. The argument here (for implicit or tacit existence of some convention, norm or rule) tends to be that given my reaction (let us say I think it is inappropriate), there must have been a convention, norm or rule implicitly or tacitly there, and this explains why I reacted the way I did. But this argument fails: first because I may have reacted differently, and second because it is my reaction (or a more thoughtful response) that can lead me to construct a convention, norm or rule (I support my attitude that pointing out there are 9 letters in the name of the station is inappropriate by saying

---

[148] Something along these lines is an actual jingle used by a radio station: tune in to 150 FM in Geneva.

that there is a norm that jingles ought to say something about what kind of music the station plays, not something as irrelevant as how many letters there are in the name).

That is one way to unravel these cases: to notice that examples of so-called tacit or implicit conventions, norms or rules often depend on that convention, norm or rule already having been constructed and used as a standard for evaluating persons, which we can notice when we look at cases where claims to there being a convention, norm or rule implicit or tacit are implausible (as with the above case). A second way in which we can unravel the cases is to change the contrast pool, e.g. in the above case of there being a convention to drive on one side of the road, we can change the contrast from one country to another by looking, within the country, at different periods in which cars are driven, such that, for example, our 'finding' might be that in such-and-such a country when cars are driven in the absence of other cars on the road, drivers drive in the middle of the road, and so it is only when there are other (or many other) cars on the road that they drive, say, on the right hand side. It would be too much to call such behaviours 'resistance actions', as if the drivers took particular pleasure from violating the convention of driving on a particular side of the road – of course, they could take such a pleasure, but the point is that their mood need not be determined by either following or violating a convention; one only thinks it must be if one assumes that the convention is always (by default) relevant, such that behaviour (on the road in this case) is always describable by reference to the convention (as if, precisely, the convention was tacitly followed by practical necessity). By changing the contrast pool, then, we can come to see what we miss when we assert, over-confidently, that such-and-such a convention governs such-and-such a community (especially when one adds that such conventions are tacit and are tacitly followed). This is not, of course, to say that it is not the case that drivers in such-and-such a community can treat a certain behaviour (driving on a certain side of the road) as not only required, but also unchangeably required – just that we cannot over-generalise the utility of this normative resource. In other words, when we make the convention tacit and we describe persons as tacitly following it, we leave no room for the contingent active treating of a convention as a guide and standard for criticism, meaning also we leave no room for the agent to act in a way where the convention is not relevant.

The point here is to show that we can be over-impressed with certain contrast cases: they make it seem as if (looking from a certain distance) it is so obvious that

such-and-such a country is governed by such-and-such a convention, and by making it seem so obvious, the case makes it difficult for us to see the limit of the claim. In making it difficult to see the limit of the claim we do not see, for instance (imagine we assert that such-and-such a culture is governed by a greeting convention that requires three kisses on each cheek): 1) how the convention could not possibly govern everyone in the culture, e.g. young children, rebellious teenagers, tourists, recent migrants, those with disabilities of various kinds (of course this list is not exhaustive; the borders of the list are porous); 2) how the applicability (the relevance) of the convention always remains ambiguous, e.g. does it apply between men, does it apply when we are greeting each other again 5 minutes after the first hello, does it apply when one of us is sweaty, etc (here the very point is that the 'etc' is indeterminate); and 3) given (2), the negotiations, sometimes awkward and hesitant, between two persons, concerning the relevance of the convention to their encounter (the point here being that although both parties can bring certain expectations to the encounter, they must be capable of suspending them or changing in the course of the interaction – the need for such suspension, change, openness to negotiation, shows that appropriateness cannot be squeezed into the following / policing of determinants).[149] No account of the normative dimension of social life can afford to ignore these ambiguities, which are missed when one focuses exclusively, or even dominantly, on how social life is governed by determinants of appropriateness and inappropriateness.

These problems are just noted briefly for now, for they are relevant to tacit knowledge claims, but we will be returning to them in the next Part, for they also concern claims about the necessity for sharing determinants.

This Part has not covered all the various possible manifestations of the governance view at the level of the mind (rather than, as was done in Part II, at the level of understanding / describing certain activities, whether it be chess playing or rationality or language use), but it is hoped that sufficient indication – including sufficient criticism – has been provided to motivate the idea that there are many important things we miss about the mind learning and working when we adopt the idea that the mind is necessarily (metaphysically or practically) governed by determinants of appropriateness and inappropriateness. The things we miss – such as tentativeness, awkwardness, hesitation, the balancing of confidence and doubt, the

---

[149] Here, then, awkwardness and hesitation is a positive thing – a sign of a second-person interaction (more on this in Part V).

exploration (including the normative exploration) of how things might be (or of what might perhaps be appropriate or inappropriate), and others – are, it turns out, vital to our participation in social life from the second-person perspective. The mind is not by default passive in the ways that the governance view often describes it: it is, instead, a mix of the active and passive – one could perhaps say, of the actively active and actively passive – and that mix is precisely what is needed to participate in social life. We shall be returning later, in Parts V and VI, to showing why such abilities and experiences are needed, and how we can describe their operation.

Part IV. Necessarily Shared:
The Third Element of the Governance View

We have arrived at the third and final component of the governance view, i.e. the argument (or more commonly, assumption) that determinants (i.e. those phenomena that are thought to be capable of determining what is appropriate or not, including what is socially appropriate) are not only necessary (the first element), and not only necessarily govern the mind (the second element), but are, in addition, necessarily shared – at least necessarily shared insofar as we are speaking at the social level or describing social action (or understanding social minds).

This third component is difficult to argue against because it is often, as noted above, not explicitly argued for, but rather posited as necessary insofar as we are speaking at the social level. In other words, the mark of the social is often taken to be the mark of what is shared (or what is agreed, what is common etc). Throughout this Part, but at greater length in Part V, we will be trying to advocate for a broader (or, perhaps better, looser) understanding of the social, i.e. one that does not require things to be shared (whether agreed or allegedly held in common in some other way), while still allowing for sharing as a contingent matter. Every attempt will be made to shift the focus away the alleged conditions for (and, as a consequence, the paths laid out in front for) interaction, and onto interactive dynamics as they unfold from the second-person perspective. Throughout this Part we will also be arguing that when we do argue or assume that determinants need to be shared (and we do this as a way of capturing the normative dimension of social life, i.e. as a matter of acquiring, accessing and acting in accordance with those determinants), then we also provide an unduly narrow, rigid, and passive picture of minds-in-interaction (the focus in the previous Part was on repercussions for understanding the mind-in-action, and thus primarily persons' relations with the environment and not with other persons).

This Part is broken down into five Chapters, and each picks up on a version or variation of the general strategy, i.e. the idea that determinants must be shared. Again, it is not the case that in all of them this strategy is made explicit; instead, for instance, in some of these versions or variations, social life is depicted as if it were a matter of learning and then re-enacting necessarily shared determinants (of what is socially appropriate and inappropriate). In each Chapter, we will look in detail at one particular view, though often also discuss surrounding, similar views, and, as we go

along, we will offer – as we have been – alternative examples or alternative readings of the examples offered, partly with a view to criticising the theory being discussed, but also partly with a view to giving glimpses of the positive picture offered in Parts V and VI.

The five Chapters are: first, a Chapter (15) on script / scenario following (in the work of Christina Bichierri and David Velleman); second, a Chapter (16) on social conventions (primarily in the work of Andrei Marmor); third, a Chapter (17) where we look at arguments for and the effect of situating descriptions of social actions within practices (especially in the work of Joseph Rouse); fourth, a related Chapter (18) to the third one, but this time where the strategy involves situating descriptions of social actions within games or contexts that are explained to be game-like (here we will look mainly at aspects of Erving Goffman's work); and fifth, and finally, a Chapter (19) that considers various versions of the claim that we act unreflectively, yet in socially appropriate or adequate ways (returning the above-analysed work of Rietveld, while also considering the work of Pierre Bourdieu and Hans Joas, as well as some recent social cognitive theories).

There are, in fact, two ways in which the material that is covered in this Part can be understood, with both of them being relevant to this thesis. In one sense, the argument in this Part concerns criticising the necessity for the social, as well as social cognition, to be conceived of as a matter of sharing (especially sharing determinants of appropriateness and inappropriateness) – in that respect, this Part is relevant because without first defeating such a view, there would be no room for the second-person perspective (which precisely does not require, though allows for, sharing). In another sense, the argument in this Part follows on from the first two Parts, for, typically, once it is argued or assumed that determinants of appropriateness and inappropriateness are necessary, and once it is also argued or assumed that determinants of appropriateness and inappropriateness govern the mind, nothing seems more natural than to extend the analysis (which is initially done at the individual level) to the social level, such that those very same determinants thought to be necessary and governing minds are now said to be shared by groups (thereby creating and thereafter maintaining, if things go well, social order). In that second sense, the analysis conducted in Parts I and II is more fundamental than in this Part; in the first sense, the analysis conducted here is just as fundamental and could stand on

its own as a specific argument concerning how to conceive of the social and social cognition.

Chapter 15. Script / Scenario Following

The idea of script / scenario -following is perhaps the most obvious starting point for an examination of the general strategy that engaging in social life (particularly insofar it has a normative dimension) involves being governed by necessarily shared determinants of what is socially appropriate and inappropriate. Perhaps the most sophisticated recent philosophical presentation of this idea is outlined in David Velleman's recent book, *How We Get Along* (2009). This will be examined shortly, but first it will be useful to set the scene with another recent account: Christina Bichierri's *The Grammar of Society: The Nature and Dynamics of Social Norms* (2006). There are, as we shall see, important differences between the two – indeed, although Velleman does not engage in detail with Bichierri's work, he does cite it (Velleman 2009, 71, fn 13) as an example of reliance on a more rigid concept (scripts) than he wants to make use of (scenarios), hence also the split in the title of this Chapter (we shall look, later, as to whether his claim that his is a more flexible concept can be justified).

A. Bichierri's Scripts

Bichierri announces at the beginning of her work that 'Norms are the language a society speaks, the embodiment of its values and collective desires, the secure guide in the uncertain lands we all traverse, the common practices that hold human groups together' (Bichierri 2006, ix). This is a lot to hope for from norms. In fact, the phrase is also somewhat misleading, for it refers to 'the language a society speaks', whereas (as her title indicates) Bichierri is more concerned with the analogy with grammar:

> I call social norms the grammar of society because, like a collection of linguistic rules that are implicit in a language and define it, social norms are implicit in the operations of society and make it what it is. Like a grammar, a system of norms specifies what is acceptable and what is not in a group. And analogously to a grammar, a system of norms is not the product of human design and planning. (Bichierri 2006, ix)

We see here, in general outline, features of views we have encountered above: e.g. the idea that what is acceptable or not (appropriate or not) is specified by some determinants (in this case, norms); the idea that what is acceptable or not, as so specified, is presented as a given to members of a group who then need to be

described as acquiring, accessing and acting (on the whole) in accordance with those norms; and the idea that (at least some of) those norms (especially in a social context) are not explicit, but that this does not mean that they are not norms – it just means that they are implicit norms (which might also mean – we shall have to see – that they need to be somehow tacitly learnt and tacitly accessed and / or complied with). But let us proceed more slowly here, and consider more of the details and examples of Bichierri's view – with particular attention to how she brings in the importance of scripts and script-following.

Bichierri's first step is to define what she means by a norm (or social norm, as she refers to it). Her theory, as she puts it, is a 'constructivist' one, i.e. 'one that explains norms in terms of the expectations and preferences of those who follow them' (Bichierri 2006, 2). Elaborated in more formal terms, Bichierri argues that norms exist where there is a 'behavioural rule R for situations of type S', and in a certain population there exist sufficient persons of that population (individuals), such that for each of those individuals, each individual 'knows that a rule R exists and applies to situations of type S', and 'prefers to conform to R in situations of type S on the condition that' they believe 'that a sufficiently large' number of persons in that population also conform to R in situations of type S and either: a) believe that a sufficiently large number of persons in that population expects that individual to R in situations of type S or b) believe that a sufficiently large number of persons in that population expects and prefers that individual to conform to R in situations of type S, and may sanction behaviour that does not conform (Bichierri 2006, 11). To this we need to add the idea that the relevant 'behavioural rule R' may 'often go against narrow self-interest' (Bichierri 2006, 2), which if we recall from Part I of the thesis, was also what quite a number of theorists argue distinguishes norms from conventions (for conventions are typically in the self-interest – narrowly conceived – of persons).

There are several ambiguities that we need to point out in the above. The first is that, on the one hand, Bichierri asserts that she offers 'a rational reconstruction of what a social norm is, not a faithful descriptive account of the real beliefs and preferences people have or of the way in which they in fact deliberate' (Bichierri 2006, 3), but, on the other hand, she spends a good deal of time (principally chapters 3 and 4) providing an account (part of which – hence our interest – involves appeal to script-following) the aim of which is to show how persons do in fact follow norms, even when they do not set out to do so, or irrespective of whether they realise it: 'my

rational reconstruction of social norms does not commit us to avow that we always engage in conscious deliberation to decide whether to follow a norm. We may follow a norm automatically and thoughtlessly' (Bichierri 2006, 3). These two claims are hard to reconcile: either the account of social norms is presented as an idealisation, or one makes the attempt to show how persons do in fact follow those norms.

This ambiguity is related to another one that runs through the book, namely: what is the relationship between norm conformity and norm following for Bichierri? In the above quotes, we referred to Bichierri saying that persons may 'follow' norms 'automatically and thoughtlessly', but if there is to be a distinction between following and conforming to a norm, it is hard to make sense of this use of 'following'. Confusion is here increased when hot on the heels of the above quotes, Bichierri refers to 'compliance' that 'is often automatic and unreflective', saying further that 'whereas the literature on social norms has traditionally stressed the deliberational side of conformity, in this book I want to emphasise its automatic component' (Bichierri 2006, 4). The reason for the ambiguity may be that Bichierri realises – despite the terms she sometimes uses – that if she wants to make the rational reconstruction view of social norms more plausible (and presumably then also less of an idealisation, and more of a faithful description), then she needs to show more than mere norm conformity. This is because mere norm conformity may reveal only that the theorist who asserts / describes some behaviour as conforming to norms finds those norms the relevant descriptive instrument for the behaviour, and this arguably tells us more about the theorist (what she thinks is required in such-and-such a community) than it does about the agents whose behaviour is being described. In other words, Bichierri needs to show that not only is there norm conformity, but there is norm conformity in the right sense, i.e. that there is norm following. The challenges that she then sets up for herself are two-fold: first, although we could say that the norms are followed deliberatively and consciously, this would greatly narrow down the scope of the view, so the need arises to account for some sense in which norm following is automatic and unreflective, unconscious and non-deliberate; and second, as we do so, the need arises for a 'solution' to the frame problem (i.e. one puts oneself in the position where one needs to say how relevance is determined for persons – with all the effects of this, which we have explored above, on how one then describes the mind learning and working).

We will consider, in a moment, how Bichierri calls upon script-following in order to offer a view of how automatic, unreflective, un-conscious, non-deliberate norm following is possible. Before we do that, let us notice another difficulty in Bichierri's general definition: recall that the definition was providing conditions of existence for social norms; recall, too, that this definition included reference to 'behavioural rules' that were applicable to 'situations type S'. This raises the following question: does Bichierri's account in fact presuppose the existence of 'behavioural rules' that are applicable in certain situations, and thus presuppose what (but under a different name, i.e. under the label of social norms) she is trying to prove that we follow? After all, where do these 'behavioural rules' come from?[150] Who has identified them as 'rules'? How do they come to be the objects of knowledge, belief, expectation and preference? Further, according to whom are those rules applicable to situations type S? It is difficult to see how one can resist the conclusion that these behavioural rules that apply to situations type S are simply presupposed by Bichierri: they reflect her understanding of certain activities or practices, or social life generally. Indeed, this conclusion is further strengthened by the thought that it is only when such rules are presupposed that the problem confronted by Bichierri arises, i.e. it is only once one has presumed the existence of such rules (and thus offers no alternative way of understanding how persons engage in social life), that one then comes to be committed to the task of explaining behaviour as conforming to them as a result of following them, which then creates the problem of how persons can be said to follow them automatically and unreflectively. The point is that one cannot restrict the scope of the importance of the rules (by restricting them, say, to deliberate and conscious guidance), for one has presupposed that engaging in certain activities and practices must be a matter of conforming to them. As soon as one opens up the possibility that engaging in the activities and practices need not be a matter of following those rules, for those rules do not exhaust what it means to engage in those activities and practices (or in social life more generally), then one also releases oneself from the requirement

---

[150] Bichierri argues that the norms come from interaction, but this can hardly be an answer when it presupposes the existence of the behavioural rules; what it suggests is that Bichierri interprets interaction as if it was producing and thereafter following those norms precisely because she presupposes the rules.

that one squeeze one's description of the mind learning and working to the learning of rules and the following of them unconsciously, automatically, etc.[151]

Let us, in any event, keep these issues in the background as we look at how Bichierri uses the idea of scripts. Social norms, argues Bichierri, 'are embedded into scripts' (Bichierri 2006, 57). 'Once a script has been activated', she continues, 'the corresponding beliefs, preferences, and behavioural rules (norms) are prompted' (Bichierri 2006, 57). Bichierri rightly thinks that one of the crucial problems here is precisely the relevance problem: how is it that the norm comes to be relevant (i.e. applicable) to a situation understood in a certain way? Again, this is crucial because in the absence of this it is difficult to see how one could describe the behaviour as norm following (rather than simply norm describable from the perspective of the theorist). Bichierri appears to acknowledge this when she argues that in the absence of evidence of salience (of how the norm appears as relevant – or applicable – to the norm follower), it is difficult (if not impossible) to assert that they are following a norm. Indeed, she says so explicitly when she refers to experiments (which we will look at in a moment) that, she says, 'unambiguously demonstrate how dependent norm avoidance is on the cues made salient in a given context' (Bichierri 2006, 59). Perhaps somewhat unwittingly, what Bichierri does is show us how rarely – in what confined circumstances we can hypothesise that – someone is following a norm, i.e. in circumstances where we (the theorists) make the norm salient to the subject (for then we can also be more confident that it is 'the same' norm for us as it is for the subject) and then we expect them to behave in certain ways (where those certain ways are what we understand to be the fulfilment of – the acting in conformity with – the norm). Bichierri tries to generalise the account to one where persons, in the course of everyday life, can be understood to see cues in certain situations (or 'contexts', as she calls them), such that those cues trigger the scripts, and because the scripts embed the norms, so the persons can be said to be norm-followers (see the diagram at Bichierri 2006, 56). But the account does not generalise, for what we do not know in these uncontrolled cases is whether, and if so how, norms become salient to persons. Again, it is only in a highly controlled environment, which we (as theorists) have set up as a context where the (same) norm is made salient to the subject and where we (again as

---

[151] To clarify, Bichierri does not argue that we always follow norms unconsciously and non-deliberatively, but just that we do sometimes or often do, such that when we do not, we follow them consciously and deliberatively: see e.g. Bichierri 2006, 5.

theorists) are also evaluating the subject's behaviour by reference to that same norm, that we can assert (and even then, this is at best a good hypothesis) that the subject is following (if he does what we expect him to do) a norm. But let us proceed more slowly and take a look at the experiments.

Bichierri refers to two kinds of experiments: field and laboratory experiments. The second is the more controlled environment, but what Bichierri means by field experiment is still fairly controlled. What tends to happen in the first kind is that a subject is 'exposed to stimuli (visual, verbal, etc) aimed at making a norm…highly available' (Bichierri 2006, 60). In the experiments Bichierri describes this norm was the norm of beneficence. The 'exposure' itself takes place in a laboratory, and what makes it a field experiment is that the subject is observed on the way home. On the way home, matters are arranged so that the subject finds 'a highly visible envelope that the experimenter has positioned in the middle of the street'; the envelope contains $20, and has no name on it. After taking the envelope, things are arranged so that the subject comes across a beggar (a confederate), and the conjecture is that 'priming a beneficence norm would significantly increase people's disposition to be charitable and generous to strangers (Bichierri 2006, 60).

There are many problems with this set up. Bichierri points to some, as when she says that 1) 'One would want to know…what subjects would do with the money they found if they had not been previously exposed to the norm' (Bichierri 2006, 60); 2) one might want a control group of subjects who find the money but receive 'neutral stimuli' to see, for instance, whether finding the money puts subjects into a good enough mood to give (some or all?) of it (Bichierri 2006, 61); and 3) that 'seeing the beggar' might have 'an effect on giving, quite independently from norm-activation' (Bichierri 2006, (61). These are significant problems, but they are not the only ones: note, first, that the experimenter in these cases is correlating a certain expected behaviour (giving money to the beggar) with following a norm (the norm of beneficence), but that there may be many different ways in which the subject's behaviour might be informed by the norm (we are following this locution here for the sake of argument) that have nothing to do with giving money to the beggar, e.g. the subject may have thought of his poor neighbour and decided to give the money to her (or the subject may have an opinion about begging that renders that an exception – to him – to being generous to others); second, we would need to know more about the stimuli, e.g. has the subject been shown pictures of giving things to beggars (if so, this

could be a case – if one gives money to the beggar – of imitating something one has recently seen?); and third, and perhaps most seriously, the norm is primed, or made artificially salient.

In laboratory settings, experiments have an opportunity to control the environment even more, but this hardly helps in the generalisation of the claims. The most sophisticated experiments that Bichierri points to here are those conducted by Cialdini et al (1990). These experiments used 'environmental cues to detect whether norms can be primed and which norms…can be primed' (Bichierri 2006, 62). These experiments are important for Bichierri because she takes them to show 'the role that situational cues play in focusing people's attention on different kinds of norms' (Bichierri 2006, 62-3). Cialdini et al worked with a distinction between 'descriptive norms' and 'social norms', where descriptive norms refer to what is 'commonly done' and social norms (also of the kind that Bichierri is interested in, for they contain 'normative beliefs': Bichierri 2006, 63) refer to what is 'commonly approved or disapproved of' (Bichierri 2006, 63). This distinction is important, for it also points to two different kinds of stimuli (used in the experiments) and also two different kinds of motivational streams (or so it was claimed): first, 'where we take the behaviour or opinions of people to convey important information about the environment, and thus imitate or adopt them'; and second, where 'we pay attention to what people approve or disapprove of' (Bichierri 2006, 64). The experiments focused on littering. The experimenters manipulated the environment in a number of different ways. They set up the environment as either clean or dirty (with trash either present or not). To this they added confederates (but only in certain versions – subjects, it should be added only participated in one version) that did various things, e.g. some, in clean environments, littered and some put litter in a trashcan. The results showed, as Bichierri reports, that 'subjects littered more in a dirty rather than a clean environment (32-40% versus 11-18%), and they littered even more in an already dirty environment if they observed a confederate littering (54%)' (Bichierri 2006, 65).[152]

On just the data above, the experiments are (obviously enough) not about 'social norms' (in the sense defined above, and in the sense that interests Bichierri).

---

[152] Bichierri also mentions that 'A smaller number (6%) littered whenever they were exposed to a confederate throwing a flyer on an otherwise clean floor', 'smaller' that is than a clean environment on its own (where littering was at 11-18% (Bichierri 2006, 65) – Bichierri does not dwell on this last case. The case is important, for it can help show that persons are more likely to do what they perceive is commonly done, including seeing some individual's actions as exceptions (i.e. persons infer that the clean environment means that that individual observed does not do what is commonly done).

Bichierri acknowledges this, for in interpreting the experiments she speaks of making 'descriptive norms' salient (see Bichierri 2006, 68). We do not want to dwell on this here, but it bears saying that referring to what is commonly done via the language of norms is confusing (on Bichierri's own definition). Put differently, what the experiments show (on the above data) is that in dirty environments, people are more likely to litter, and in clean environments, they are less likely to, i.e. they are most likely to do what they might see as the commonly done thing in such-and-such an environment. This could be a simple case of imitating what others do, and it seems to confuse things to refer to 'norms' here (of course, we are not attributing this confusion to Bichierri, for she is merely reporting Cialdini et al's experiments).

But putting this aside, the experiments become more relevant where they introduce two further conditions, meant precisely to indicate 'normative influence'. The two conditions are: first, in one condition, 'subjects observed a confederate picking up a piece of litter from the ground. Regardless of the state of the environment, almost all refrained from littering' (Bichierri 2006, 69) – this was also distinguished from cases where the confederate was seen throwing his own handbill into a trash can (the second is said to show personal disapproval, whereas the first is said to indicate a normative expectation possibly connected to sanctions); second, in the other condition, the handbills that subjects were given (to throw away) had messages on them, some 'priming a norm' (e.g. urging recycling) and others not (they included pleas to vote), with the result that subjects littered less in the case of the recycling message. The way Cialdini et al presented this data, and the way that Bichierri reads it, is that they show that 'focusing subjects on social norms always produced the expected effect [not littering, or littering less], irrespective of the prevailing descriptive norm' (Bichierri 2006, 70). This is important for Bichierri, because it connects up with how she wants to introduce scripts (as embedding norms), the idea being the scripts are activated by certain stimuli. They are also important in her argument for sharing norms, for she says: 'A stimulus…has to be interpreted within a context of background (collective) knowledge to be effective in activating norm-abiding behaviour. If no littering norm existed, or if there were no shared recognition that the situation is one to which the norm applies, neither the confederate's actions nor the appeals to recycling would have had any consequence on behaviour' (Bichierri 2006, 70).

The way the data is being interpreted here and the way in which the notion of 'norm' is used are both problematic. Take, first, the case of the confederate picking up a piece of paper from the ground and putting it in a bin. It is true that on observing this we may conclude that the confederate cares about whether there is rubbish on the floor there or not. But this hardly shows that a 'norm' is activated the subject when the subject refrains from littering. For instance, it might show that the subject refrains from littering because – having just seen the confederate pick something up – the subject would be scared to hurt the confederate's feelings (for the subject doing precisely the opposite of what the confederate just did would show the subject has no respect for the confederate's efforts), or scared to do something that would potentially cause conflict between the subject and the confederate (it would damage, we might say, the quality of the interaction between them). In other words, what the subject may be said to be sensitive to here (when the subject does not litter in this situation) is the other person's likely reactions, and not a social norm (the subject might be said to care about those reactions, and not the norm). To see this, one can take cases where we are not already disposed to think there is a norm (after all, we have all seen signs 'do not litter' etc, so we all too readily think of a 'norm' in this scenario): imagine we are fruit-picking, and I see you not only not dropping rotten fruit on the ground, but also picking up rotten fruit left behind into your apron pockets; I may very well do the same, but I am hardly necessarily doing so because your behaviour primes a norm for me, but just because I respect that you care about something, and out of respect for you (and not the norm) I do something that would convey disrespect to you (as dropping the rotten fruit would).

Take, now, second, the case of the messages on the handbills. The problem with this is that it can be thought of in two related ways: first, it presupposes what it is trying to prove, i.e. the existence of a norm, for the messages are precisely about not littering (saying you ought to recycle is arguably another way of saying that you ought to place your litter in the appropriate bin) – here, the problem is that the stimuli is the norm, so it is a self-perpetuating prophecy: to stimulate the norm is to activate the norm, but what Bichierri needs if she is to generalise (as she wants to generalise) the account is stimuli that does not itself contain the norm; second, the 'cues' that Bichierri needs for her account cannot be the same as that which the cues are supposed to activate, for to make them the same is to say that persons are simply asked to reproduce the same behaviour – there is no generality to the norm, and thus

there is no norm; in other words, what Bichierri wants to think of as a norm (and thus something 'more general' that 'prescribe[s] or proscribe[s] an entire class of actions in a variety of situations', Bichierri 2006, 70) is not what the stimuli is: the stimuli is basically priming (and then evaluating the subject on the basis of) a certain expected action.

The general point we can make, then, is that Bichierri's account shows that what she is interested in is a kind of social engineering where it is assumed that certain specific kinds of actions (putting litter in garbage cans) is 'pro-social', and that the aim should be to try to encourage (or better, manipulate so as to encourage) persons to reproduce those very same actions. She says as much when she says that we ought to be 'altering the environment in such a way that individuals will find certain desirable norms salient and act on them' (Bichierri 2006, 63), and the point being made here is that 'norms' are a misnomer here, for what Bichierri has in mind is the reproduction of certain specific actions which she assumes are always going to be 'pro-social'. This may work in a controlled environment, as in these experiments, where what is assumed to be pro-social (and thus also the content of the norm) is a certain specific action, and all that is relevant to the evaluation of the subject is whether or not they perform that action. But the difficulty with this is that it presents social life – in its normative dimension – as if it were a matter of mechanically reproducing the same actions (which are assumed to be the right / appropriate ones), i.e. it presents social life as if there is a set of situations we can impose in advance, where each situation places paths in front of persons and tries to force them into performing precisely those actions and no others. To see the difficulties with this one needs only remind oneself of the beggar case: social life, including in its normative dimension, is not such that persons are confronted with a 'beggar situation' where the right action is to give the beggar something (and where self-styled designers of social life ought to be manipulating the environment so as to make that the salient thing to do for the person in that situation). As is indicated by the example of the person who thinks of his poor neighbour and does not give the beggar anything for that reason, or perhaps does not even consider the beggar's pleas, there are many ways in which we might exercise what we might call our ability to care about the plight of others, or care about others' vulnerability. To present social life as if it were a matter of one situation after another, where in each situation there is a right thing to do and a wrong thing, and to try to engineer social life so as to make us do that thing, is to neglect and

put in jeopardy the richness of our abilities to care (not only about the vulnerability of others, but about many other things – including learning to care about something we had not previously thought was deserving of care).

It is against this background that Bichierri introduces scripts. Scripts are a kind of 'schemata', where 'schemata' are defined as 'cognitive structures that represent stored knowledge about people, events and roles' (Bichierri 2006, 93). The schemata that Bichierri is interested in are those 'that describe appropriate sequences of events in well-known situations', e.g. 'descriptions of what happens at restaurants, soccer games, theatres, and lectures' (Bichierri 2006, 93). These schemata (following Schank and Abelson, 1977) are scripts (Bichierri 2006, 94), and it is in these scripts that she argues that social norms are embedded. We have already seen that the first step here is that of 'categori[sing] a situation as being of a certain type', for it is on the basis of this categorisation that – according to Bichierri – a certain 'script is activated that will involve players' interlocking roles, a shared understanding of what is supposed to happen, and even prescriptions for unexpected occurrences' (Bichierri 2006, 94).

We should be beginning to see how the necessity for sharing comes into the picture: learning to become a member of the community is acquiring knowledge of the same scripts, and the same cues (the 'categorisations of situations of a certain type') for scripts, such that when we are faced with a situation of that type, we categorise it as such a type, and then because the script is triggered we interact with others (who have similarly categorised the situation, and thus been similarly triggered) according to the script. We need scripts, says Bichierri, because it is only thanks to scripts that we can 'know what to expect of each other' in the absence of being 'acquainted with each other of know[ing] of each other's past performance' (Bichierri 2006, 96). It is these scripts that 'guide us' (all of us, at least within that community) 'in everyday life', and, furthermore, given that 'we tend to perceive scripted interactions as "right" and appropriate' (Bichierri 2006, 96), we might say that it is thanks to, and in terms of, our sharing of scripts that we can explain the normative dimension of social life. Having pictured matters in this manner, the door is open for Bichierri to argue that both our learning and our reliance (our application) of these scripts (which includes the categories with which we interpret situations as of a type that trigger the scripts) is automatic, unreflective, non-deliberative and unconscious: 'What we have…is implicit, nondeclarative knowledge: We are unable to tell how we learned our categories or articulate our present knowledge clearly and

precisely' (Bichierri 2006, 97). Here, and in a way that may remind us of the Chapter on tacit knowledge (Chapter 14), Bichierri speaks of the implicit knowledge in bicycling or driving a car, or learning to recognise a Mozart piece (Bichierri 2006, 97). 'Norm following', she concludes 'is similar in this respect to bicycling, or the ability to recognise a piece of music: Once a schema is activated, we tend to follow the norm by default, without being able to tell what prompted it or which features of the situations acquired particular relevance' (Bichierri 2006, 97). Here is her overall summary:

> Interpreting and understanding a social situation involves comparing it to similar ones we have encountered in the past. What we view as similar will determine which category will be retrieved and which script or schema applied. Social norms, being embedded into scripts, are activated as part of a process that also triggers the beliefs and expectations that support and justify conformity. The cognitive process of organising and structuring new stimuli through retrieving from memory similar situations is thus responsible for framing an interaction as one in which fairness or reciprocity norms are activated. (Bichierri 2006, 98)

The basic idea (which we have seen before in this thesis) is that we learn specific things, which we are then able to recognise; what is added to this here is that this act of recognition triggers certain scripts, which function like rules for interacting with others. This only works if it is assumed that others share the same scripts, and although some room is made for there being 'ambiguity' in how the situation is to be interpreted, and thus also 'negotiation' of what situation it is (see Bichierri 2006, 98), it is still assumed that there is a certain stock of situations, which all persons who interact in that community share, such that in order to interact they need to come together to interpret the situation in the same way, so that that interpretation can trigger the same script (which then enables interaction).

What needs to be asked here is this: can we, and ought we, assume that 1) there must be a certain stock of script-triggering-situations (identified by the behavioural rules referred to earlier) that are, as it were, constitutive for that community; and 2) persons must have learnt the same scripts and the same categories for classifying situations? Further, is the only (and indeed the most important) ambiguity we experience in social life one where we run through (without realising it) the stock of 'script-triggering-situation' knowledge we have in order to see which script applies, and then negotiate with others (presumably tacitly, though Bichierri

also says, unsurprisingly, that matters are aided here by subjects discussing which script to follow: Bichierri 2006, 98) as to what kind of situation the one in which we found ourselves is? The problem here is not so much with saying that it is possible that we (both you and I) interpret our encounter as a certain type of situation that then triggers the 'same' script for us; rather, the problem is with offering this as a description of what goes on whenever (or most of the time) we interact, and thus as an exhaustive, or at least dominant, picture of the normative dimension of social life (with repercussions for how we picture minds-in-interaction learning and working).

Perhaps we get a clue for how to unravel this picture in the first sentence of Bichierri's summary above. There, Bichierri suggests that at bottom lies a capacity to evaluate things as similar, or sufficiently similar, and that it is this capacity that 'determines which category will be retrieved', and that in turn determines which script will be applied. This is an interesting way of putting it, for it is crucial what we begin the picture with: do we begin it with a presumed set of situations that contain such-and-such features that persons passively detect (with determined consequences), or do we begin with an ability to evaluate that is itself active, and not a matter of detecting but (potentially, if the need arises) constructing situations? Notice that if we begin with the second – with the ability – then we allow for the possibility that we may construct a situation as one that is familiar to us, and familiar in the sense of that triggers a certain script, but (and this is crucial) we do not present this possibility as a necessity, but rather allow for the possibility that we may construct a situation that is unfamiliar to us, probing in a more hesitant manner for ways to interact with others (and in response to their actions) that we have not 'acted out' or 'carried out' before.

Part of the issue here is how we picture ourselves as learning: do we picture ourselves as learning something specific that we are doomed only to recognise, or do we picture ourselves as learning something more inherently flexible and creative? The above quote that points to there being a stage of evaluating similarities suggests Bichierri may be open to the suggestion that we learn something inherently more flexible and creative, but the difficulty is that when she discusses categorisation, and certainly in the examples (we have been discussing) of situations-triggering-scripts, she seems to begin with a presupposed stock of such situations, and thus describes persons as learning – and then recognising, and eventually re-enacting that which the act of recognition triggers – some specific series of actions. Thus, for instance, although when she is describing categorisation she begins with the promising idea

(taken from Rosch 1978) that categories are 'collections of instances that have a family resemblance' (Bichierri 2006, 83) – this is a promising idea precisely because it pictures us as learning something more flexible, for it is not a matter of 'recognising' some specific properties of objects, but rather rougher matching, or messy overlapping between images – she goes on to describe such processes of learning in the following way:

> As I am exposed to a series of examples of a category, I learn about each individual example, but I also learn about the items as a group. I might be aware that I have learned about a specific example, but I might not be aware that I have acquired category-level knowledge. The latter kind of knowledge only emerges when I am confronted with new examples of the category and have to classify them. For example, I may be aware that I also learned something about the pieces as a group, or what is common to all of them. When I hear a new Mozart piece, however, I will be able to recognise it, even if I am unable to describe the features that make it a Mozart piece. (Bichierri 2006, 84)

Notice that last phrase: 'features that make it a Mozart piece'. This is crucial, for it suggests that she presupposes that Mozart pieces are such pieces in virtue of certain properties (the case is set up in such a way that it is presupposed that the new piece of music I am hearing is a Mozart one). The effect of this is to reduce what we learn to the learning of those properties, and to reduce the act of seeing this as a Mozart piece to recognising those properties. This is precisely not the Rosch picture, the very point of which was to disassociate the perfect matching of properties to classes, and to introduce more activity into the role of the mind, which must exercise an ability (never determined, but open) to construct similarities.[153]

Notice, however, that even this more flexible notion of learning categories does not go far enough: where it falls short is that it still pictures persons as having to see what we encounter as something we have previously experienced (as either sufficiently like this, or sufficient like this), and thus unable to create a new category. It is an improvement on mere recognition of properties, but it does not go far enough. In any event, in Bichierri, we do not get past the initial (non-Rosch-like-account) of perfect matching, or of us learning stocks of categories that we then recognise in situations so as to trigger specific scripts.

---

[153] To say 'construct similarities' is, in fact, like saying that we create relevance, for we actively decide what should be a similarity and what should be a difference.

Recall where we began with Bichierri's account: the idea that norms existed where there were 'behavioural rules R for situations of type S', and there were populations such that there existed individuals in those populations (a sufficient number of them) where each individual 'knows that a rule R exists and applies to situations of type S', and 'prefers to conform to R in situations of type S on the condition that' they believe 'that a sufficiently large' number of persons in that population also conform to R in situations of type S and either: a) believe that a sufficiently large number of persons in that population expects that individual to R in situations of type S or b) believe that a sufficiently large number of persons in that population expects and prefers that individual to conform to R in situations of type S, and may sanction behaviour that does not conform (Bichierri 2006, 11). We have seen how Bichierri has tried to make this rational reconstruction plausible, but here we again must ask: is her account one that *can* make the above reconstruction plausible? Recall, from a moment ago, what Bichierri said about implicit learning and the implicit exercise of knowledge: that we can learn to recognise certain situations as of a certain type and then automatically and unreflectively do so in certain instances. This seems incompatible with the idea (in the above reconstruction) that persons have certain beliefs, including quite complex normative expectations (for the expectations are not just about what some specific person will do, but that there are sufficient members of the population that would do so, or sufficient number that would not only do so but also sanction those that do not do so). The point here is that in order to make it the case that Bichierri's account of how we follow norms automatically and unreflectively can make plausible her rational reconstruction, we must basically presuppose the existence of the behavioural rules; only when we presuppose that existence can we try to give an account of how persons, as a result of being members of that population that is governed by those rules, come to learn them and follow them automatically and unreflectively (for on that basis we impute, as it were, certain beliefs and expectations to them). But, of course, this presupposes what it sets out to explain, i.e. that there are such norms (it just looks as if there is no presupposition because Bichierri refers to 'behavioural rules' instead of norms).

Earlier in our discussion of Bichierri, we quoted her as suggesting that scripts were necessary because otherwise we would need to rely on our knowledge of the person in question, or on some past history with that person. It is important to resist this dichotomy between script-following on the one hand, and some demanding

knowledge / history with the particular person with whom we are interacting. In other words, putting things this way misleadingly suggests that it is only by way of scripts that we can generally interact with (unknown) others, and this fails to make room for the possibility that we may be exercising abilities to interact with (unknown) others (including respecting what we may observe them as caring about) without script-following. Script-following is not our only option; indeed, as we have tried to argue here, it cannot be our only option – or, more accurately, it can be only if we presuppose, as we are not entitled to presuppose, that social life is made of script-triggering-situations, and that persons (insofar as they are capable of engaging in social life so understood) are programmed into detecting that they are encountering such-and-such a situation, which triggers such-and-such pre-determined paths for interaction. This not only misrepresents social life (makes it more rigid than it could possibly be), but is also greatly narrows our understanding of minds-in-interaction learning and working.

B. Velleman's Scenarios

It was mentioned at the outset of this Chapter that Velleman's picture in *How We Get Along* (2009) shares some affinities with Bichierri's, but that it also differs in some respects from it. Indeed, we mentioned that Velleman explicitly distances himself from the notion of a script (which, like Bichierri, he associates with the work of Schank and Abelson), and prefers the concept of a scenario instead (claiming that it is more flexible). Before we can get to his use of that concept, however, we need to work through two prior stages: Velleman's understanding of what he calls 'acting' and his understanding of what he calls 'valuing' (though, somewhat confusingly, he calls the chapter where he deals with this – chapter 2 – 'reacting'). Velleman's book is rich with insight, and we will here be focusing on selected aspects of his view (primarily the first three chapters, and thus the chapters that deal with or support his account of interaction – his later chapters are in a sense applications of the first three to the idea of morality). Given this selection, it is important to note that we are not evaluating Velleman's project as a whole (nor indeed, the project of this latest book in his entire impressive oeuvre).

Velleman's notion of acting depends on a contrast with behaving, where acting is something that is raised to that level by an extra element (something that is allegedly missing from mere behaviour). He illustrates this early on his chapter on

acting (chapter 1) by referring to the different ways in which we can cry. There are three such ways, though the first two are extremes (and the way that he is interested in is the third, which sits in the middle). In the first extreme way of crying we cry with an 'involuntary outpouring of emotion' – 'bursting into tears at the first shock of pain' – and in the second extreme way we contrive our crying, as a baby does when it continues to cry long after the pain has subsided (Velleman 2009, 10). In between these two ways of crying lies a third: a mature, responsive, adult-like crying, which Velleman describes as an 'authentic but also voluntary behaviour of the adult who, after absorbing the first shock, indulges in a good cry' (Velleman 2009, 10). This person can 'tear at his clothes a bit', but not too much (Velleman 2009, 11). Velleman then – and this is a crucial move – connects this mature, responsible, controlled crying to crying according to one's conception of crying: 'A person who actively cries does not just let some behavioural program take over: he has an idea of what he's doing, and he brings the manifestations of his emotion into accord with that idea. His emotion initially propelled him blindly, but now it carries him through the witting fulfilment of an action-concept that he has in mind' (Velleman 2009, 11).

At this point, Velleman introduces an analogy that he draws on throughout the book (and to a very large extent in the context of interaction), i.e. an analogy with acting (or interacting) out a role, or playing (for instance, improvising) a character. In this specific context of action, this allows him to make the following move. When actors improvise a character, they must focus on making the sounds and gestures that they need to make in order to be acting (improvising); their focus – or attention – is on making those sounds and gestures; at the same time, however, they must be communicating themselves to the audience in such a way that the audience will see the performance as a performance of that character – the performance, then, must make sense as a performance of that character; it must be intelligible as a performance of that character. Now if the actor is focusing on the sounds and gestures, then he is clearly not able to focus – to attend to – the sense in which his performance is intelligibly (to himself and to his audience) one of that character; nevertheless – or so Velleman argues – there is a sense in which that actor is 'aware' of what it would make sense for the character to do (see Velleman 2009, 23).

Velleman sets things up in this way for two reasons: first, because he wants there to be a layer above the mere making of sounds and gestures, or mere behaviour, and he takes it that elevating the mere behaviour to action requires (as we have seen)

some conception being in place;[154] and second, because he does not want to make this account so implausible so as to require that a person first decide what he wants the character to do and then does it, for this would essentially rid the performance of any 'improvisation'. Velleman, then, wants to have both rationality (as he himself calls it, i.e. he wants to present our actions as choices made in light of a rationale of when / how the action makes sense: see Velleman 2009, 18) and improvisation (in the sense of not requiring a deliberational sequence, as when we first figure out what would be the rational thing to do and then execute it: see Velleman 2009, 21).

The two relevant questions here are: first, why does he want both; and second, does he / can he succeed in having both? In a way, we have already answered the first question: Velleman wants rationality (of a kind of implicit or tacit rational layer on top – he speaks here of 'tacit self-attributions' of what would make sense: Velleman 2009, 21) because he wants to distinguish between mere behaviour and something more, i.e. action – but of course to say that is to push the problem back one step, not give an answer, for we can ask: but why does Velleman want a distinction between mere behaviour and action?

That is a question we have encountered often in this thesis, and it seems to be connected to the idea that we want to present the action as appropriate or capable of being evaluated as appropriate or not. Velleman does not make this his ostensible aim, and he reverses (as he himself puts it – we will discuss this in more detail in a moment when looking at valuing) the order of explanation, saying that what is intelligible is what explains what is appropriate (rather than the other way round) – but his ultimate aim is still this sense of wanting to show that something can be and either is or is not appropriate. This move is accompanied by the idea that something being appropriate or not must be appropriate or not by reference to something that determines whether it is, but seeing that it would be implausible to think that we access and process that something every time we perform, Velleman pushes that something into a tacit layer above, something that we are aware of but do not attend to.

We have there the motivation for the distinction and for the positing of the tacit layer, but is the view viable? In trying to make more plausible this idea of an

---

[154] He speaks sometimes of second-order dispositions connected to the character the actor is playing, as distinct from the first-order dispositions to make sounds and gestures (Velleman 2009, 15); so the view is that the actor is enacting the second-order dispositions he has, and doing so in a way that is accompanied by awareness, but not by attention.

implicit ground for action (of acting for a reason – as he sometimes puts it – but without being initiated, in any self-conscious way, by a reason), Velleman sometimes refers to a 'disposition' that gets revealed in a possible counter-factual, i.e. we are aware of it as making sense because if we were challenged, we would make sense of it, or it would make sense to us (Velleman 2009, 25).[155] The notion of a disposition of this kind suggests a period of time, and indeed Velleman does seem to think that over time, we can we can come to shape our first-order impulses (to make sounds and gestures) in a way that we would be disposed to see as whatever we act out as intelligible (as making sense).

There are two things to say in response: first, it is not clear that introducing a temporal element helps, for it does not seem necessary that even if we have a disposition to find certain of our actions (or a character's actions) intelligible, that we are determined to find them intelligible in just the way we previously have (perhaps it makes it likely, but it is too strong to say it determines it); and second, as we have previously argued in this thesis, the counter-factual cannot be depended on to reveal anything about why we did what we did (if there is a why – and of course part of the point here is to show we can resist implicitly inscribing a why), because how we respond to being challenged, and thus how we articulate our reasons for doing something, is a context with its own pragmatic pressures that cannot just be assumed to have unproblematic access to why we did what we did.[156]

Towards the end of the chapter on acting, Velleman says something quite interesting, which we may be able to build on: he says that we should not think that the kind of awareness of intelligibility he is pointing to (a kind of self-understanding, at least where we imagine we are an actor performing to ourselves as a tacit audience) is 'making sense *of* oneself as one is' but 'also a matter of making sense *to* oneself, perhaps by being otherwise' (Velleman 2009, 32; original emphasis). Here, Velleman suggests something more active for the mind, and also as a result something more open normatively, but he does not develop this further.[157] In fact, he goes on to curtail it by saying that 'although a particular agent would make sense to himself after turning over a new leaf', it may be that 'he cannot make sense of taking the steps

---

[155] His exact words here are: 'he can still be aware of it and of what *would* make sense in light of it' (Velleman 2009, 25; emphasis added).

[156] This idea is elaborated on, together with examples, in Section 10.D.

[157] Later, when speaking of rationality in chapter 4, he mentions 'rational ambivalence' in positive terms, but he does not develop this and focuses instead on 'rational authenticity': see Velleman 2009, 90.

required to turn it', suggesting that the latter is a case of reaching a 'rational dead end' (Velleman 2009, 33). He further comments on this by saying that 'Aristotle was right to emphasise the importance of a good upbringing: a good upbringing is, among other things, an induction into a self-concept that won't lead into rational dead ends' (Velleman 2009, 33). The idea here of rational dead ends is interesting, but not for the reasons Velleman suggests; indeed, he paints the idea negatively. What we need to do – if we want to build on Velleman's notion of 'making sense *to* oneself, perhaps by being otherwise' is make room for the possibility – indeed desirability – of experiences where we do not (already) know what would be better for us, what would make sense for us to do. Velleman presents this as a kind of practical inertia, and in a sense he is right: this is a case precisely where we are not guided by some conception of what it would make sense for us to do, precisely because we are not sure; an opportunity for 'being otherwise' presents itself to us or, better, we create it (here, we as it were fashion a new character for ourselves to improvise on), and we feel our way either towards it or away from it, the point being that nothing determines (nothing drives, or guides, explicitly or tacitly) whether we go one way or the other. The point we want to stress here is that this is a positive thing, and not something to be marginalised and presented in a negative light; it is both valuable and vital for the exercise of rationality, and not its dead end.

It may seem that this is difficult idea to make sense of – for in a way we are making room here for something contingent and arbitrary, something that does not immediately make sense to us – but this is precisely what we need to do if we are not to succumb to the kind of view where we only ever do that which already makes sense to us (or *would* make sense). We need to open up – rather than close – what *might* make sense to us (and if we do that, following Velleman's direction of explanation, we would also open up what we might appreciate as appropriate), but to do this we need to resist the idea that there is a layer on top, something tacitly existing above us, something which we are tacitly aware of, that as it were gives our sounds and gestures a rationale, a ground, a basis on which it is always justifiable.[158]

---

[158] Perhaps we can put this more colourfully: sometimes, even as adults, we burst into tears and sob uncontrollably and sometimes we contrive our feelings (perhaps sometimes even without meaning to); these moments in which we are very vulnerable, and not in control, are important, and they can change us and transform us in various ways: change – or being otherwise – is not always, as Velleman would want it, a matter of rational, controlled choice (again, we can have the experienced of controlled crying, but to make it control of that kind a necessary element of any action is to over-generalise the

More light can be shed on these issues if we go on and look at how Velleman explains 'valuing'. Velleman's analysis begins in promising terms, for he makes a distinction between judging *stricto senso* and valuing – this suggests he may be making room for something akin to what we have been calling in this thesis the context of appreciation. Where his account of valuing differs from what the context of appreciation seeks to promote is that he pictures it as he does action (or, better, improvised action), i.e. as a matter of being regulated, in the wings, by 'what makes something valuable' (as we shall see in a moment, this is partly subjective and partly objective). The context of appreciation, as it is understood in this thesis, is different because it does not posit a tacit regulator in the wings, i.e. it does not require the necessary presence of a criterion according to which that which we appreciate as appropriate (or not) really is appropriate or not (for oneself, or objectively). It does not posit such a tacit regulator because it alleviates itself of the need – or anxiety – of thinking that 1) appropriateness and inappropriateness must be determined; and 2) that there are determinants that can determine what is appropriate and inappropriate. Consider the following passage:

> To find someone likeable or admirable or enviable, to find something interesting or amusing or disgusting – these are what might be called guided responses, responses that are somehow sensitive to indications of their own appropriateness. Guided responses are not value judgements, since they are still conative or affective, rather than cognitive attitudes. But they resemble judgements in being regulated for appropriateness, and so they are more than mere responses. Finding someone likeable is more judgemental than merely liking him, but it need not entail passing judgement on his likeability. It is rather a matter of liking him in a way that is sensitive to what makes him worth liking. We can mark the partial similarity of such guided responses to value judgements, or evaluations, by describing them as instances of valuing. (Velleman 2009, 35-6)

The key in this passage is in the second last sentence: 'a matter of liking him in a way that is sensitive to what makes him worth liking.' It is this that positions us in the passive state of being confined to detecting – or to developing over time a sensitivity so as to be able to detect – what is (by some pre-existing criterion) really likeable (or admirable, disgusting, etc). Notice, too, that Velleman refers to 'merely liking': by this he seems to mean something like a mere reaction, i.e. an automatic response.

---

importance of control). But we mention this only as an aside, for crying may not be a good case to have in mind.

Is what we have been calling the context of appreciation 'merely liking'? No, because 'merely liking' is a category invented to contrast with an act of evaluation that is necessarily (though not necessarily explicitly) guided by something (i.e. something that makes that thing likeable, admirable, or disgusting, etc). It is imagined by Velleman as a kind of reflex, something completely mindless – like the uncontrolled burst into tears we met in the context of action. We need not, of course, dismiss the presence of such reflexes or responses, but there is room for something in between a reflex / response and a guided valuing, and that space can be conveniently thought of as occupied by what we have been calling the context of appreciation (it is that space that allows us to be active, and not always passive, vis-à-vis the normative).

There are many similarities between Velleman's treatment of valuing and his treatment of acting. As with acting, when we value something we are controlling our reactions by filtering them through (though not consciously or deliberatively in the sense of attending to, but nevertheless remaining aware of) what would make sense for us: here, in the context of valuing, it is not what would make sense for us to behave, but what 'would make sense for him to feel' (Velleman 2009, 39-40). Put succinctly, 'the considerations that make something valuable, by providing reasons to value it, are considerations in light of which valuing it makes sense' (Velleman 2009, 40), and it is those considerations to which we become sensitive whenever we engage in valuing.

The reason why Velleman refers to things (or considerations) that make something valuable is to get around a familiar problem, i.e. the trap of thinking that whatever a subject admires is what makes something admirable; as he says: 'the conditions of appropriateness for a response cannot be read off the actual responses of the relevant sensibility. What's appropriate to admire isn't merely what admiring subjects actually do admire' (Velleman 2009, 41). We have already had occasion to reflect on this thought many times: the point to make about it is that it invites a split into the subjective and objective when one thinks that 1) what is appropriate or not must be determined; and 2) one neglects to make room for both the possibility that when one is valuing one can be a) treating some criteria as unchangeable; and b) at the same time be capable of ignoring, changing or simply considering the criteria to be irrelevant and unhelpful and thus creating some way in which one (and others after one) may find that thing valuable in a new way (or they may not – there is a risk). Velleman is right that we cannot always value without external constraints, but where

he goes wrong is that he thinks that something must always and already be valuable in respect of something, such that all that the person does is develop a sensibility or sensitivity (or 'attunement', as he also sometimes calls it) to what makes that thing valuable. Put another way, Velleman is looking for a perfect fit between what makes something valuable and our abilities to find things valuable, but this necessitates presupposing and then fixing how something can be found valuable. The result is the untenable positioning of the mind as perpetually passive vis-à-vis that which can emerge as normatively relevant.

Velleman makes a lot of room for the subjective, but as a result of maintaining the split between the subjective and objective, which itself is the result of thinking there must be a 'fundamental standard of appropriateness' (Velleman 2009, 42), he turns into something overly passive: thus, for instance, he makes room for the idea that the considerations that make something valuable for us may not be the considerations that everyone shares, but that they may be considerations that make sense for us – would fit into a coherent web of our various forms of valuing (i.e. fit into, say, a web of ways in which I admire things, ways in which I find things disgusting, etc). For Velleman, whenever we value we are under rational pressure to make sure that the way in which value something is sensitive to how it would fit into that web of ways in which we have valued things, e.g. I am under rational pressure (which I am aware of, but do not need to – though I can – attend to) to explain to myself in the course of controlling my finding something funny whether it makes sense for me to laugh at it, where what would make sense is both a matter of how I have laughed at certain jokes previously, but also how my finding such-and-such a thing funny fits in to my other ways of valuing (e.g. finding things amusing, admirable etc). My overall valuing sensibility, then, is 'cultivated under rational pressure to be responsive to recognisable kinds of things, which constitute what is admirable, inspiring, or disgusting for me' (Velleman 2009, 46).

We can see, in this, how Velleman is attempting to put intelligibility at the foundation: what makes something valuable to me, and thus how I value it, is what makes sense for me to value. Velleman articulates this in very subtle and impressive ways, and makes it very plausible, but the problem is still with the way he generalises it to all cases of valuing, such that anything that falls short of valuing in his sense – where it is regulated, in the wings, by this rational pressure for that act of valuing to make sense of my overall practice of valuing – cannot be valuing. This creates an

overly rigid and passive view of how I can (sometimes, not always) value things (appreciating them to be admirable, funny, inspiring etc). Velleman is right that I can be aware of how my finding something funny fits in to a certain image I might try to construct of my self (as a character, say, in the play of my life), and he is right, too, that I can come, at some later stage and not in light of some particular immanent act of valuing, to try to organise, make coherent, unify etc my ways of valuing (on this, see Velleman 2009, 42), but he does not give us good reason to think (in fact, we have good reason not to accept the idea) that we must always and can only value in this careful, responsible manner (or that valuing properly so-called – elevated above 'mere reactions' – must be so careful and responsible).

As noted above, although Velleman does make a great deal of room for the subjective shades of valuing, he also has an objective story. He does this on the basis that although differences exist among us, we have a 'shared human nature' (Velleman 2009, 44). There are, he says, 'fixed points of human nature' (Velleman 2009, 42), such that some things are desirable / admirable etc for any human being, because 'desiring them will make sense for anyone; other things simply cannot be desirable because desiring them won't make sense for anyone…. Everyone would make more sense to himself desiring' certain things and not other things, 'given his natural endowment as a member of the species' (Velleman 2009, 44). His examples of shared human nature are the following: 'an array of physiological appetites; an aversion to pain, separation, and frustration; an inclination toward pleasure, connection, and the fluid exercise of skill; the inborn and automatic flight-or-flight response; an interest in the human face and form; an initial dislike of snakes, spiders, blood, and the dark; and so on' (Velleman 2009, 45).

It is difficult to evaluate this proposal, partly because Velleman himself qualifies the above list by saying they are 'nearly universal' (Velleman 2009, 44 – how can they be *nearly* universal when they are also said to be 'fixed points' of 'shared human nature'?), and partly because, when he comes to apply them, he says that these 'fixed points of human nature place different degrees of constraint on the intelligibility of different responses' (Velleman 2009, 44 – in other words, we would need to know more about what these 'different degrees' are). It is one thing to say – as Velleman says at one point – that there is a 'universalising tendency', i.e. a tendency amongst members of a population (and also two populations merging) to converge on finding things valuable in the same way (Velleman 2009, 47), but it is

another to posit some allegedly universal features of shared human nature that precede and determine (? – though in 'different degrees'?) any particular acts of valuing or any convergence in valuing amongst a group.

Given his emphasis on coherence, you might ask whether, and if so how, Velleman deals with inconsistency in our ways of valuing: do these, according to him, never occur? His answer here is a little ambivalent (his discussion is at Velleman 2009, 48-58). At first, Velleman appears to want to say that ways of valuing can conflict in a person by that person valuing the same event in different ways. However, he then goes on to say that what appears as a conflict is not really a conflict, and this for two reasons: first, that we only think there is a conflict because we fail to notice that the valuings are attached to the same event but under two different descriptions (in other words, their object is not really the same event at all); or two, because the two valuings occur at two different times, and occurring at two different times may mean, for instance, that they are either particular or general (e.g. I can find unfavourable, because it makes sense to me to find it unfavourable, a 14 year-old mother giving birth, but after she does actually give birth, I may greatly value the particular child she gives birth – here, as a result of occurring at two different times, the valuing is either particular or general, and thus there is no conflict). This makes it difficult to see how Velleman thinks there could be conflicts; it looks as if what might first appear as a conflict is but a misdiagnosis; this is a way of presupposing that there is a rational order, and we must just detect it, rather than allowing for a period of conflict (or ambivalence) that we then constructively resolve in some way unfamiliar to us (for example, by characterising our valuings as having a different object, or occurring at different levels of generality, etc). In other words, the absence of real conflict or inconsistency in Velleman's account is another sign of presenting the mind as overly passive (precisely because it is positioned passively vis-à-vis the normative – in this case, the coherence of valuings).

Let us, now, turn finally to Velleman's account of interacting. Insofar as we interact, according to Velleman, we need to see each other as fellow self-enacting improvisers (Velleman 2009, 59), but ones also committed to making each other intelligible to each other. The process is a rather complicated one, and perhaps can best be understood by Velleman's image of the two of us constituting a third-person perspective on ourselves, and thus the two of us serving as our own audience to our interaction (Velleman 2009, 60). The first crucial move here is that Velleman's

approach imposes a certain task on our interaction, i.e. the task of understanding each other. Further, the way he understands this task is a matter of me communicating to you what would make sense to me, or at least opening myself up for you to understand what would make sense for me, and vice versa. The construction of an imaginary audience is designed to make us want to unify the way in which our interaction could make sense to both of us, so that we interact (and as we shall see in a moment, we interact appropriately) only insofar as – and indeed thanks to – our interacting making sense to the two of us (to that imaginary audience, that third-person, of which we are members). As with acting and valuing, none of this needs to happen in a way that is mediated by attention (although, working in the wings, there is awareness – in this context, some kind of shared awareness).

We can imagine that insofar as we are trying to make ourselves understood, one way in which could do this would be to imagine ourselves as two actors on a stage, playing two parts, and trying (without attending) to make this performance an intelligible one to ourselves, but the question we can raise (at this stage) is: why think that whenever we interact we need – or we even want – to make each other understood, let alone make our joint performance understandable (i.e. perform in such a way that we would find it understandable – or intelligible – if we were to stop and think about it)? There are two ways to go here: one is to say that when we walk, dance, or debate or discuss together it is possible to perform these tasks in such a way so as to control how we do so in order to make the process intelligible and our actions understandable to each other, but that it is also possible for us to be unsure, hesitant, groping around somewhat blindly, for we do not quite know where we are going or how to get there; another way to go is to say that both are part of the same process, i.e. that there is always some sense in which, insofar as we interact or dance / debate together,[159] we move in and out of balance, attempting sometimes to share and sometimes receding into a private realm, moving towards the other but also distancing ourselves. We need not resolve this problem at the moment: at this stage, we just want to flag the possibility that it may be too much to think that interaction is necessarily a matter of both of us perpetually striving to be understood and wanting to make our performance an understandable one (to the imaginary audience of us).

---

[159] There are differences between interacting and doing something together, but let us put these aside for the moment.

But let us assume, for the moment, that we want to make ourselves understood to each other, and that we want (without attending to wanting) is to make our interaction intelligible to ourselves. How do we do so? For a start, Velleman takes it that we are both self-enacting improvisers (i.e. actors in the sense in which he developing the idea of acting). This helps, for it means that we encounter each other as persons who already have a certain self-understanding (a coherent self, if you like), i.e. a self that can be intelligible to another. As Velleman puts it: 'My drive towards self-understanding favours making myself amenable to an interpretation that won't ramify exponentially as I come into contact with others' (Velleman 2009, 66). We can see how important it was for Velleman to begin with the notion of acting: we begin with selves who, prior to interaction, already have cared about their own intelligibility (to themselves), such that now, in the encounter with one another, the way we could be understood is limited, and limited precisely by the ways in which have made ourselves intelligible to ourselves (recall that when we act, we are acting in ways that would make sense to us; this is what limits – what makes possible – how we could appear as intelligible to another). In the abstract, this seems fine, but where it becomes problematic is where we ask: but how can you access how I would make (based on how I have in the past made) myself intelligible to myself? It is difficult enough making oneself intelligible to oneself, but how could others have access to how I make myself intelligible to myself? Velleman does not so much resolve this dilemma, as he proposes something that facilitates our making ourselves intelligible to each other, namely our sharing a certain limited range of possibilities of how one can be intelligible (both to oneself and to others, or intelligible together). This becomes necessary because otherwise the question of access looms too large; after all, if we begin with selves who have made themselves intelligible to themselves, and then go on to interact with others, then one needs a theory of how persons can access each other. One can get around this problem if one says that insofar as we interact, we need not so much access to each other's ways of making ourselves intelligible to ourselves, but rather the ability to find each other intelligible by acting on the same stage. How, then, can we act on the same stage? Here is Velleman's answer: by sharing stocks of ways of playing characters on a stage, i.e. precisely scenarios (as Velleman calls them, in preference over scripts).

In a way, Velleman wants both things, i.e. he wants us both to try to understand the other as the other is intelligible to themselves, but he also wants to

make our performance intelligible by acting on the same stage. The point being made here is that there is a tension in these two things that Velleman does not quite resolve or address. One way to avoid this tension is to say that neither do we need access to how others find themselves intelligible, nor do we, in order to interact (partly because we need not make our performance intelligible to the imaginary audience of us) need to share stocks of scenarios – in other words, the tension may be created precisely by presenting acting and interacting in light of the alleged need for making ourselves intelligible to ourselves, to others, and to each other simultaneously.

This tension is important to analyse in a little more detail, so let us attempt articulating the above in another way. Velleman begins, as noted above, with solitary agents: 'A solitary rational agent would be under similar pressure' – i.e. to act in a way that would be intelligible to himself – 'even if he ever interacted with agents'; over time, this agent would 'develop a way of doing things, which would then become self-reinforcing, because doing things that way would begin to make sense, not only for the reasons that initially made it intelligible, but also for the additional reason that he has developed standing knowledge of its intelligibility, which has entrenched it in his way of life' (Velleman 2009, 77). Two such solitary agents can then come into contact with one another. Here is where Velleman has two stories about what might happen (the two that are in tension). First, when they do so, 'they would begin figuring out one another's way of life', which 'would entail not only noticing one another's way of doing things but also interpreting one another's modes of self understanding' (Velleman 2009, 77). But second, 'there would also be convergence for the sake of social self-understanding' (Velleman 2009, 77). When he explains this 'convergence' in this context, Velleman speaks of it coming after the two agents have figured each other out, i.e. 'When an agent figures out how the other understands himself, he will learn categories under which he can be fairly confident of making himself understood to the other' (Velleman 2009, 77). Elsewhere, however, as we shall see in a moment, Velleman describes one agent figuring out the other by reference to what scenarios (for interaction) the other's behaviour might be triggering, and, indeed, it is difficult to see what other access to the other's modes of self-understanding Velleman has in mind. The one example he provides is that of adopting 'the other's way of doing something because it instrumentally superior: maybe the other agent has invented a better gizmo for trapping mice' (Velleman 2009, 77), but

the invention of a gizmo for trapping mice hardly reads like the other's mode of self-understanding.

The obvious question is how – since he requires it – does Velleman suggest one agent has access to the other's self-understanding? Notice that it does help to say that the agent can observe the other's behaviour, for observing the other's behaviour does not give access to the other's self-understanding. Velleman does not take this further, but instead enlists – as noted above – the assistance of scenarios: the other's behaviour is interpreted as potentially triggering such-and-such or such-and-such a scenario; for instance, he says: 'without a shared vocabulary, we would struggle to find understandable ways of responding to others who haven't made themselves understood to us' (Velleman 2009, 80 – but here Velleman does not say what this other way of others making themselves understood to us may consist in). Indeed, he refers to us interacting without understanding each other as a 'Babel', and suggests that we can 'escape from this Babel by developing a common pool of situationally defined action concepts, and by drawing on this pool for our ideas of intelligible things to do' (Velleman 2009, 80). But by the time this is said, we have lost sight of the other, and our attempt to understand them. In other words, to call this process of encountering others, and not understanding them, a Babel, is to avoid, rather than confront, the relationship to others *as others*, and to paper over the difficulties by saying that – for the sake of social understanding – we treat others as but characters in a scenario that is familiar to us, and that we assume is also familiar to the other.

There are two things to say about this: first, we can agree with Velleman that sometimes when we interact with others we treat them as if they share some understanding of the situation with us; second, we can assert that acknowledging this first point does not mean that we need to lose sight of the experience of an encounter with an other whom we do not understand, and who we feel does not understand us (the uncertainty, hesitation and tentativeness in our relationship with others is important and valuable – and not a Babel; to call it a Babel is to impose a misleading dichotomy between a certain, valuable order and an uncertain, destructive chaos). It is crucial, then: 1) to make room for this sense of tentativeness in relationships with others; and 2) to consider what might happen in it, i.e. not to assume that it must result in one of us actually succeeding in understanding the other (figuring out the other, as Velleman puts it), but to consider the possibility that a) we may not come to understand one another (and that is ok) and b) part of our relationship with others, as

others, is that we cannot think that we have 'figured them out' or 'succeeded finally in understanding them' (they must *remain other* to us, for if they do not we are no longer interacting with them, but just treating them, from a third-person perspective, as a character in a play familiar to us).

The fact that this tension exists in Velleman's analysis, and that he skirts over the sense in which we may encounter an 'other' that we do not immediately figure out or understand thanks to some allegedly common scheme for understanding, suggests there may be problems with how the entire dilemma is set up. Indeed, there are such problems: they include the idea that the self presents itself to others as something available for being figured out, i.e. as so stable that it can be accessed by another; and the idea, connected to that one, that the other can be 'figured out', as if the other's modes of self-understanding were somehow transparently observable in their actions (for their actions would be parts of scenarios that they must, on Velleman's view, share with us if they are to interact with us). Instead of this way of seeing it, it might be better to say the following: first, we do not need to 'figure' each other out in order to interact with another ('figure out' is simply too strong – it suggests detecting something there awaiting to be recognised, and thus positions us too passively, too conservatively, with respect to how we can appreciate others; it also positions others as passive objects, awaiting being figured out); second, the lack of this need to 'figure out the other' does not mean that we cannot have encounters with others where we have the experience of not understanding them (and not being understood; in other words, we can have an experience of wanting to understand an other and the other remaining unintelligible to us); and third, these encounters where we feel the 'otherness of the other' (an otherness we do not immediately relate to and that is unintelligible to us) are important – we need them precisely in order to interact with others, rather than simply to imagine them, from the safe distance of a third-person stance, as characters in scenarios familiar to us. There is more to say about the issues (especially the encounter with otherness, which is looked at in detail in Chapter 23), but let us put this aside for a moment – we will come back to it when we consider interaction from the second-person point of view.

In the same way that Velleman aims for unity in acting (the unity of the self), so he aims for unity in interaction, or what he calls 'social harmony' (Velleman 2009, 69). This has several effects. The first is that he pictures persons tending 'to gravitate towards others who view' themselves as they view themselves (Velleman 2009, 66).

The second is that it presents encounters with others as either operating in contexts where we (somehow already) share the manner in which we can find our performance intelligible, or we where we seek agreement on some manner in which we could find our performance intelligible.[160] The same problems arise here that arose for acting: are conflicts at all possible (or are they only ever misdiagnoses); and, when, and if so how, can new scenarios / roles arise (or is 'creativity' only ever a matter of creating a version of, or an improvisation on, a pre-existing scenario / role)? The question of conflict in the social context is a two-fold question: first, does Velleman make room for encounters between persons who do not make themselves intelligible to each other; and second, how does he see such encounters working – what happens in them, e.g. do they need to be resolved so that the two persons / groups must converge (e.g. by one conforming to the other)? This is another way of asking: how much unity / harmony does Velleman either presuppose or think that must there be?[161]

This is where we can reach into Velleman's discussion of scenarios. There is one piece of background here that is important: the reliance on scenarios is part of a picture of social cognition, but in the background is the issue of cognition generally. We have seen how Velleman thinks about this to some extent already – through his treatment of acting – but in this chapter on interacting he also adds an idea that reveals quite a lot about how he understands cognition generally: 'the fundamental form of understanding', he says, 'is generalisation, the "grasping together" of disparate particulars under general principles, whose utility for comprehension is compromised when they are multiplied, or larded with exceptions, qualifications and special conditions' (Velleman 2009, 81). At the social level, this has the following effect / aim: 'To eliminate social distinctions where possible usually results in a more comprehensible way of life' (Velleman 2009, 81). What matters – at either the individual or group level – is that life is comprehensible, understandable, intelligible, and this means that we are always striving (even without realising it) to generalise, universalise, unify, simplify, and thus iron out or 'eliminate' as Velleman puts it,

---

[160] When he discusses this, Velleman seems to presuppose the existence of a certain limited range of such manners, e.g. of a limited range of roles we could play – in other words, he allows for 'negotiation' and 'experimentation', but this always seems to be negotiation as to which out of a limited number of scenarios to enact, and experimentation in the sense of trying on different – but still choosing out of a limited range – roles to play: for negotiation, see Velleman 2009, 67; for experimentation, see Velleman 2009, 160.

[161] Notice that this predilection for unity, sameness, etc appears also in the idea that whenever we interact we try to understand each other, or make ourselves intelligible, or seek agreement.

'exceptions, qualifications, and special conditions.' Velleman, then, puts incredible pressure on convergence, conformity, harmony, unity, order – on things being the same.

There are many things we can say in reply to this: do we need such unity for the comprehensibility of life (whether the comprehensibility of my life to myself, or life with others)? Could it be that a person's life, and our life together, can be comprehensible without unity – or with occasional unity contingently built on top of an underlying and contested diversity? Why assume that conflict and contestation is a bad thing for comprehensibility? But even if one enlarges comprehensibility to be an indeterminate mixture of sameness and difference, of falling in and out of harmony, of both controlled and accidental change, does one need to think that I / we always have to seek comprehensibility (without realising it)? Velleman seems to present a picture of social life that is either always and already harmonious, or always tending towards harmony (where any sense of disharmony is immediately quelled), and he seems to do this by presenting persons as not only careful and responsible to themselves (i.e. always doing what makes sense to them), but also as interacting together in ways that are careful and responsible, i.e. enacting performances that make sense to themselves as imaginary audiences of us. But is this not an implausible utopia (in fact, really a nightmare), or at the very least an over-generalisation of the value of unity / harmony, including an over-generalisation of the importance of coherence / simplification for cognition?[162]

Let us see how this plays itself out in Velleman's appeal to scenarios. Before he gets to the term 'scenario' Velleman points to the notion of 'scripts' in Schank and Abelson (Velleman 2009, 70 – it is also in this context that he cites Bichierri 2006). In a footnote, Velleman notes that the term ('script') may have been badly chosen by Schank and Abelson, for it implies 'that actions and utterances are mandated with more specificity than Schank and Abelson actually have in mind' (Velleman 2009, 70, fn. 13). Because of this connation with specificity, Velleman prefers them term scenario, which 'suggests a greater degree of indeterminacy, leaving room for improvisation' (Velleman 2009, 70, fn. 13). Let us look, then, to see what Velleman

---

[162] As an aside we can say that at stake here is a certain attitude with which we might theorise about phenomena, i.e. either an attitude that places emphasis on unity, order and sameness, or one that acknowledges the importance of these, but makes room for – and not in a way that merely sees as negative – diversity, change and difference. It will be no surprise to the reader to learn that this thesis favours diversity, change and difference, while nevertheless striving to make room for, and see the value in, the contingent active treating of things as unified, ordered and the same.

means by improvisation, and thus how much indeterminacy his account suggests. First, Velleman cites the following example from Schank and Abelson:

> Schank and Abelson's favourite example is the restaurant scenario, which can be pictured as a tree diagram of how a visit to a restaurant typically unfolds. Either you wait to be seated – in which case, you may or may not be asked whether you have a reservation – or you are permitted to seat yourself; then you wait until someone brings a menu or tells you what's available, or both; then you and your companions take turns telling that person your choices from the available items, leaving out condiments, which are already on the table, and dessert, which is ordered later; and so on. Even if you knew that restaurants are places to eat, you would have trouble extracting a meal from one of them if you didn't know how the scenario goes. If you didn't know the scenario, of course, you might ask for directions at the door; but you would then have to know the 'asking for directions' scenario. (Velleman 2009, 71)

Velleman does not critique this example; in fact, critiquing it would not have made sense, for Velleman has already noted that he thinks Schank and Abelson's account is in general sound (just that they used an unfortunate term for it). How much indeterminacy, however, is in the above case? Where is the improvisation?[163] What Velleman seems to present as indeterminacy, then, is basically a matter of choosing one out of the few fixed possibilities on offer: we detect which scenario is being triggered (is this a restaurant that takes bookings, if so wait till you are seated, etc); there is no sense in which we can find our own way – we are always presented as reliant on scenarios (consider the last sentence: even when we don't know the scenario, we need the scenario for asking what the scenario is). The above 'tree diagram' may remind us of rather mechanical action – a matter of computing fixed choices (if X, then Y) – and, indeed, Velleman refers to (and seems to endorse) Schank and Abelson's argument 'that a robot would need to know many such scenarios in order to simulate an intelligent agent' (Velleman 2009, 70).[164] There is room here for robots, and plenty of determinacy, but not much – if any – for indeterminacy and improvisation.

What does Velleman mean by 'improvisation'? We get our best clue form the following passage:

---

[163] We shall not ask how this is less specific than 'scripts', for Velleman does not give us a contrast case, but it is difficult to imagine what could be much more specific than the above restaurant scenario.
[164] Velleman says nothing here about how such an agent would decide which scenario was relevant – but of course this also ignores the bigger problem, namely the assumption that we cannot create our own possibilities for action, but that we depend on the environment being laid out or organised for us in terms of fixed possibilities that we then choose between.

Vast stretches of our social life are governed by conventional scenarios… We are like an improvisational theatre troupe that, over many years of performing together, has developed an extensive repertoire of scenes that any member of the troupe can initiate in the expectation that the others will follow suit. Again, we do not face one another on a bare stage, where we must converge on a scenario chosen arbitrarily from our repertoire. We face one another in determinate circumstances, with more or less determinate identities that have been formed in prior negotiations. But even if you are unmistakably my colleague, and you knock on what is unmistakably my office door, a question remains whether you are about to initiate the philosophical-query scenario, the unfinished-committee-business scenario, the invitation-to-lunch-scenario… In any case, you have already initiated the knock-on-the-door scenario, according to which, after I say 'Come in', the next move is yours. I await your move, from which I will then take my cue. (Velleman 2009, 77)

Improvisation, then, seems to be a matter of choosing amongst possible choices: our past experience has stocked us with scenarios, and when we confront a situation, we run through various possibilities, until we finally 'detect' which scenario is being triggered by the other's behaviour. The 'indeterminacy' that Velleman seems to be referring to is marked by that process of waiting, i.e. of figuring out which of the scenarios we know is being triggered. As we become more experienced – as we learn more of these scenarios (though it is not explained how we do so) – the indeterminacy decreases, and we develop the kind of sensitivity that can tell immediately (or at least very quickly) which of the available scenarios is being triggered.

Ought we to be convinced by this picture? This thesis argues that we ought not. The first step in distancing ourselves from it is to see that indeterminacy can apply to the situation, i.e. put more clearly, we are not confronted – and ought not to be described as being confronted – with a situation that contains certain possibilities for action; indeterminacy does not apply merely to choice among possibilities, but also to what those possibilities are. This is because we are not passive when it comes to the possibilities for action: it is we who may actively think that some possibilities (perhaps ones we have relied on before) are also relevant here, and it is also we who may actively create possibilities – not by standing back, as it were, and deciding what to do (though this can happen too), but by creating possibilities in the course of action or interaction as it unfolds. This second possibility may not be easy to see: it demands that we see the body not as the locus of habits (acquired, for instance, as a result of internalised rules or norms, or internalised conceptions of how to do things), but as

the locus of something much more inherently creative and flexible; the body, in other words, can orient itself by exploring, by groping, in the course of being hesitant and uncertain – the body need not be that which is used by the mind (even tacitly) as an instrument for reaching certain (pre-given) ends. This creation of possibilities is, in fact, best understood in interaction from the second-person perspective, the idea being that it is not me who must always impose the possibilities (either by treating familiar ones as relevant, or by creating new ones), but that what possibilities arise for me in the course of interaction can be created by you. The point is that I still need to be active in taking these invitations or opportunities up (in being open to seeing them – being open to being invited or being given an opportunity that is unfamiliar to me), and this is why I am not describable as but a passive detector of possibilities previously familiar to me.

Notice that as soon as we do not presuppose – set up the case so that it seems incontrovertible that there is – a set of possibilities for action that is (somehow) inherent in the situation, we also make room for the idea that when we arrive in a new place, a restaurant say, we can feel our way hesitantly and tentatively to appreciating how others might be finding what we do; the important thing is that a large part – the most important, foundational part – of relating to each normatively is not re-enacting always and already shared scripts and thus being assured of acting appropriately (on this understanding of appropriateness), but rather relating to each other as others, and thus being precisely responsive to ways in which others might find what we do appropriate or inappropriate in ways that we do not share. Velleman's view of the normative dimension of social life is one where there is always an underlying social order of appropriateness and inappropriateness in place, such that what we must do is conform to that – if we do not, we will be outlawed: 'If upon entering a restaurant you departed too far from the relevant scenario, you would becoming a kind of social outlaw, and like an outlaw you would find yourself in a no-man's-land where others would have to take unorthodox measures to deal with you' (Velleman 2009, 78). In other words, Velleman sets things up as if it were a black-and-white, on-off, matter: either you follow the paths determined by the scenarios in this place, or you will be treated with the ominous sounding 'unorthodox' measures. This treats the individual as someone unable to relate to others normatively unless they know what the required paths are, and others as unable to relate to that individual normatively in any other way than as close-minded police officers who monitor that individual for conformity

to what they are necessarily committed to seeing as appropriate and inappropriate. The effect of this is to squeeze out any possibility for interaction; all forms of interaction, on this view, are either (successful) re-enactments of a pre-existing social order or (unsuccessful) insults bound to lead to beatings.[165]

It is possible that, as with valuing, what Velleman is concerned with – what pushes him to set up social interaction in such a rigid way – is that he fears that in the absence of such an account, he will end up with mere subjectivity, i.e. each person imposing their own 'scenarios', such that we have a confused Babel of individual voices and no 'social harmony' whatsoever. We get a clue that this is what concerns him when he says: 'What we generally cannot do…is strike out entirely on our own, improvising without any shared scenario at all' (Velleman 2009, 87). Velleman is right that it is no alternative to say that social interaction is a matter of the chaotic, random bumping into each other of isolated monads (a kind of Babel), but he is wrong to think that that – either Babel or social harmony, either disorder or order – is the dichotomy facing us. What social interaction needs, as we will see in detail in Part V, is a mix of balance and imbalance, where both balance and imbalance is positive in its own way. We 'muddle through' (as Velleman himself puts it, Velleman 2009, 1), but this does not mean we always and necessarily treat each other as actors in familiar scenarios (though we may, on occasion, treat each other in this way). Instead, we are perfectly capable of relating to each other normatively in ways that create unpredictable possibilities for interaction, controlled by neither of us, where we remain different to each other, and where conflict, contestation, uncertainty come and go, without needing to be resolved. What we need, at bottom, at least as revealed by the second-person perspective, is not shared knowledge of scenarios, but an ability and willingness to relate with one another where precisely we are not sure who the other is or what we will do together.

We will stop here with our analysis of script / scenario following. Many of the difficulties we have confronted are also ones that we will encounter (together with some others) in the Chapters that follow – hence also beginning with the ground covered by Bichierri and Velleman. In the next Chapter, we consider quite a different account: Marmor's analysis of social conventions.

---

[165] Here we encounter, once again, the story of Khushim, which we met in the epigraph and introduction to this thesis.

Chapter 16. Social Conventions

Back in Chapter 1 of this thesis, it was promised that we would return to a recent attempt to offer a theory of social conventions distinct from (and allegedly an improvement over) those offered by Margaret Gilbert and David Lewis (which we very briefly described above). The attempt promised was that of Andrei Marmor's recent *Social Conventions* (2009). The focus here will be on Marmor's positive proposal, and we will be examining it with a view to seeing whether Marmor gives us reason to think that insofar as we speak of the normative dimension at the social level, we ought to assume or posit the existence of shared determinants of what is socially appropriate and inappropriate.[166]

A. Conventionality

Let us begin with Marmor's definition of 'conventionality':

> A rule, R, is conventional, if and only if all the following conditions obtain:
> 1. There is a group of people, a population, P, that normally follow R in circumstances C.
> 2. There is a reason, or a combination of reasons, call it A, for members of P to follow R in circumstances C.
> 3. There is at least one potential rule, S, that if members of P had actually followed in circumstances C, then A would have been a sufficient reason for members of P to follow S instead of R in circumstances C, and at least partly because S is the rule generally followed instead of R. The rules R and S are such that it is impossible (or pointless) to comply with both of them concomitantly in circumstances C.
> (Marmor 2009, 2)

This is a general definition, and we will see later that Marmor introduces not only kinds of conventions – e.g. co-ordination conventions (as familiar from Lewis) as well as what Marmor calls 'constitutive conventions' – but also what we might classify as degrees of transparency of conventions (here Marmor makes a distinction between 'surface conventions' and 'deep conventions', and he places particular emphasis on the latter).

One preliminary matter concerning terminology needs to be observed. The general definition above refers not to conventions but to rules that are said to be conventional. By 'rule' Marmor says he means 'the content of a linguistic form',

---

[166] An earlier, and considerably different, version of what follows appeared as Del Mar 2011a.

which can be 'valid or correct irrespective of practice' (Marmor does not develop what he means by valid or correct). This general definition of a rule becomes – as we shall see – somewhat difficult when one begins to speak of conventions with low degree of transparency (what Marmor calls 'deep conventions'), for it is difficult to see how rules that have linguistic form are as deep, or as non-transparent, as Marmor presents them to be. Marmor also does not trouble himself with distinctions between conventions, norms and rules. He tells us, for instance, early on in the book that 'conventions' are 'species of norms', and then adds: 'they are rules that regulate human conduct' (Marmor 2009, x). Later, again, Marmor defines 'norms' as 'rules' that are 'followed by a population or, at least, …regarded as binding by a population' (Marmor 2009, 3, fn. 1). The result is a classification of conventions as kinds of norms, and norms as kinds of rules (sometimes also referred to 'social rules': Marmor 2009, 3). From the perspective of this thesis, keeping in mind the distinctions introduced in Part I, this usage is unsatisfactory.[167] However, to keep matters simple, and in the name of reading as generously as possible, we shall try to put these terminological issues aside, and focus on Marmor's analysis of how communities are allegedly governed by conventions.

The most important feature to note of the first condition (in the above general definition) is that it requires persons in some population to 'follow' rules. Marmor uses this term 'advisedly', in that he posits a distinction between conforming to and following rules, and argues that following a rule requires regarding the rule as binding. Marmor is not always clear by what he means by this, i.e. what it means to 'regard a rule as binding', and the closest he comes to an account is one (which we have encountered many times before in this thesis) that depends on a counter-factual proof, i.e. he says, 'at least upon reflection, people would say that they behave in a certain way because the relevant conduct is required by the convention' (Marmor 2009, 3). We have had occasion to criticise the idea that any such reflective justification or report can unproblematically provide proof of following a rule, so we shall not repeat this criticism here (see Section 10.D.).

Consider, now, the second condition. Here, Marmor argues that 'there is a reason, or a combination of reasons, call it A, for members of P to follow R in

---

[167] One imagines that this terminology would also be resisted by many contemporary theorists of conventions, who may of course acknowledge that conventions, norms and rules have something in common, while nevertheless insisting that they are not explicable in terms of one another.

circumstances C.' It is not easy to follow Marmor's analysis. On the one hand, he says that the conventionality of the rule does not depend on what reasons people follow for following the rule: people may follow the rule for all kinds of reasons. At the same time, however, he wants to say that there is a reason, or a combination of reasons, underlying the rule, i.e. the reason(s) for following it. He says as much when he says that the reason derives from what makes the action required by the rule valuable (Marmor 2009, 5), and when he adds: 'it is not part of this condition of conventionality that members of P must be aware of the reason, A, to follow R' (Marmor 2009, 5). The question we may pose here is this: who is identifying 'the reason' or 'the combination of reasons' for following the rule? Marmor mentions the following example, which can help us understand this condition: 'there are some Orthodox Jewish communities who believe that Hebrew is a holy language descended directly from God. In fact, they only speak Hebrew in religious contexts, and use Yiddish for everyday life. Surely Hebrew remains conventional (to the extent it is, of course), even when spoken by those Orthodox Jews' (Marmor 2009, 5-6). The example suggests that what Marmor wants to say (in this condition) is that whatever reasons people follow for following the rule (in this case, whether or not they think that the reason for speaking Hebrew is that it is a holy language) does not affect the conventionality of the rule. But if this is what Marmor is saying, then this seems to undermine the condition as it is phrased: for if the reasons for which persons follow the rule do not affect the conventionality of the rule, why do we need a condition that says that there are reasons for following it? To rephrase: what Marmor appears to be saying is that what is not important for the conventionality of the rule is whether persons think of it as conventional or not (i.e. whether the reasons for which they follow the rule are ones that would ordinarily point it to being conventional, e.g. one of a number of possible languages to use or non-conventional, e.g. the one, holy language), but if that is what he saying it is not clear why there needs to be a condition for the rule being conventional that there be a reason, or reasons, for following the rule (it is also not clear on what grounds Marmor can simply impute a reason 'underlying' the convention).

Part of Marmor's discussion of this condition seems to go back and forth between two things: 1) following a reason for doing something (where doing something happens to be what is required by the rule), and 2) following a rule and being potentially aware of following it as a rule (of regarding it as binding). Consider

the following: 'The point here is that to the extent conventions are rules, and to the extent that conformity to a convention is an instance of following a norm, there is always the *potential* for awareness that in complying with a convention one follows a rule. But again, what the reasons for the rule are, or what kind of rule it is, is not something that the agent must be aware of' (Marmor 2009, 7; original emphasis). The trick here is to figure out what Marmor thinks the agent must be aware of. He seems to want to say that what is important for the conventionality of a rule is that people follow it, in the sense that they could potentially justify their behaviour in light of it (where it assumed that their 'behaviour' is what the rule requires, and where it is taken for granted that persons have access to, and are being honest in pointing to, the reason for their behaviour).[168] But here, what persons point to is not some reason (e.g. whether the rule is treated by them as holy, or as one of many possible rules that they follow because otherwise are following it), but precisely the rule; the rule is the reason. If that is so, we again reach the same conclusion: it is difficult to see why Marmor wants to say – as this second condition says – that there are reasons for following the rule, and that people follow those reasons without realising that they do so ('the reasons', as he says, 'for following a convention do not have to be transparent': Marmor 2009, 6).

Perhaps the best we can say here is that Marmor's condition should be understood in the following way: when the rule is conventional, people following it (when they regard it as binding) means that they need not be aware of its status, i.e. it does not matter (the rule does not become non-conventional) because some (or even all) of the people following it do so for a reason that would otherwise normally indicate that the rule was non-conventional (e.g. they say that they speak Hebrew because it is a holy language). But then read this way, the condition presupposes what it seeks to explain: it presupposes that the rule is conventional. It is as if Marmor wants to say this: 1) what is important is that people follow the rule (in the sense of doing what it requires and regarding it as binding as they do so); 2) it is not important what reasons they have for following it; 3) but it is important that they can be described as following a conventional rule – described as if their reason for following it was because it was conventional (i.e. that they can be taken to be treating the rule as

---

[168] See Section 10.D. for doubts about this approach to 'reports'.

conventional even if they report otherwise). What is not clear is on what basis Marmor can make the third step.

It is as if Marmor wants to appear as if he is taking the attitudes of the persons he is describing seriously, but nevertheless ignore their attitudes for the purpose of establishing what he thinks (on some other grounds, really) is a convention. We need to keep track of this tension as we proceed (we will see that it arises again in the context of constitutive and deep conventions).

Consider, now, the third condition: 'There is at least one potential rule, S, that if members of P had actually followed in circumstances C, then A would have been a sufficient reason for members of P to follow S instead of R in circumstances C, and at least partly because S is the rule generally followed instead of R. The rules R and S are such that it is impossible (or pointless) to comply with both of them concomitantly in circumstances C.' This third condition is designed to capture the arbitrariness of (conventional) rules. Earlier, in Part I of this thesis, we defined conventions as like norms in the sense that they depended on persons regularly performing the behaviour that is the object of the convention or norm as well as regularly exercising the relevant attitude, but not like norms in that they tended to be in the interests of persons to regularly follow them and were regularly followed because persons observed others to be following them. The way Marmor sets things up, a rule can be said to be conventional when it is arbitrary, i.e. when there is at least one another potential rule, such that if persons followed that alternative, they would have had sufficient reason to do so partly because others did so. The counter-factual language should alert us to some difficulty being skirted over: it is as if, again, Marmor presupposes the existence of what is conventional, and then describes people's attitudes in terms of that presupposition. That he does so is suggested by the very fact that he only refers to *potential* attitudes. In other words, Marmor wants to appear as if he is taking attitudes seriously, but really what is doing is: 1) positing some rule as conventional; and 2) saying that it is conventional because there are alternatives that could be followed but are not. The point being made here is that he is not grounding these claims in the attitudes of persons.

Notice that the same move is made with respect to whether persons are following rules (which is another way of saying whether a rule can be used to describe their behaviour): Marmor appears to take attitudes seriously when he says that persons must regard the rule as binding, but this is immediately thereafter hidden

under a mere possibility (a potential report or justification). Further, as we saw in the second condition, the status of the rule is presented as independent of the attitudes of persons. Matters are exacerbated here in the third condition because of the very idea of arbitrariness, i.e. the very idea seems to require that persons be aware of alternatives, but nevertheless treat the behaviour required (by the conventional rule) as the one to perform because others perform it (the other alternatives might be just as good, but this is one is chosen for the reason that others follow it). But putting things this way is not open to Marmor because he does not want (for reasons we need to still analyse) make the existence of the rules (as conventions) dependent on persons' attitudes to them (in a way, he already said so in conditions 1 and 2 – in 1, by saying that all that is needed is potential awareness, and in 2 by basically arguing that it does not matter what reasons people have when they follow the rule). The same move is made here in the third condition: the possibility of an alternative that, it is asserted, *would* be an alternative, is said to be enough to make the rule conventional (and the potentiality of attitudes is used to make this more psychologically plausible).

This becomes clearer when we look at the examples Marmor uses:

> Consider, for example, the case of a greeting convention. Suppose that there is a reason to greet acquaintances' in some conventional manner. Now let us assume, for the sake of simplicity, that there are only two possible alternatives: we can either greet each other by shaking hands, or else by just nodding our head. Presumably, some people would prefer the hand-shaking option, while others (perhaps because they are not so keen on physical contact) would prefer nodding… As long as the reason to act in concern with others is stronger than people's preference for one of the options, whichever rule evolves as the common practice is likely to be followed. And it would be arbitrary in the sense defined here. (Marmor 2009, 9)

What makes this example interesting is the following crucial question: is the conventionality of the rule dependent on persons (or even a person in a group) having encountered some alternative (perhaps as a result of an encounter with some other group, e.g. a group that shake hands rather than nod heads) and having decided, nevertheless, to stick with what members of the group typically do (because they typically do it) despite the fact that there may be room for preferring (or being neutral) about the encountered alternative as to how to greet others? Notice that if we did so restrict the idea of conventions, we would be limiting their explanatory scope, for they would become dependent on – in a way, they would consist in – the attitudes

of (on our view, contingently, not by default) treating them as conventional. Marmor does not want to do this, and in order to avoid doing it he has to make us accept several things: first, that some other alternative just is an alternative to greeting others (that nodding heads and shaking hands is equivalent, or sufficiently equivalent); second, that it does not matter whether that alternative is actually encountered, but just that it could be; third, that if it were encountered, persons would react by noticing it but nevertheless not following it, and thus keeping their original way of doing things; and fourth, that they would react that way to a potential alternative (assumed to be an alternative) shows that their original way of doing things always was a convention.

There are many problems with these assumptions: first, there are no alternatives without the attitude of members of the group treating them as alternatives (to think there are is to put the cart before the horse: to describe something as being of a certain nature prior to persons conceiving of it as such);[169] second, there is no sense in positing a potential encounter with an alternative, partly because there is no sense in positing an alternative (persons must treat it as an alternative), but partly because – and this connects with the third assumption – it is not pre-determined how persons would react to the encounter with something they treat as an alternative; third, if persons' reactions re not pre-determined – they might, for instance, adopt nodding, and do so on grounds that nodding is far superior to hand-shaking (say because it is more hygienic, and thus not just another alternative) – then it also follows that it is not the case that prior to the attitude persons take what they do has the status of being conventional.

There is another underlying difficulty here. It has to do with how we describe cases (how, in other words, we say 'what' it is that persons 'typically' or 'conventionally' do). At what point do members of the group come to think that 1) they engage in the practice of 'greeting' (that that is, as it were, the higher level, the more general, category); and 2) that their 'mode of greeting' is hand-shaking? Marmor does not confront these questions: he just describes the members of the group as engaging in the practice of greeting and as engaging in that particular mode of it

---

[169] Of course, one could argue for this, but then one has to come clean about it (in a way that Marmor does not). Further, is it plausible to argue for conventions being by nature conventions, without any need for an attitude? Is it not, at the very least, less plausible to argue for conventions being by nature conventions than water being by nature water? In any event, the point is that Marmor does not offer arguments – his analysis reveals he is already committed to seeing certain behaviours as conventional.

(i.e. hand-shaking). In doing so, he imposes a particular way of describing what the members of the group do. But what makes him so confident that his way of describing what the members of the group do is an accurate or relevant one? Here, one should answer: there is nothing to make him accurate or relevant, the point being merely that he does not consider the possibility that his description is one of many possible descriptions (he privileges his description, but gives us no grounds upon which we can agree with how he does so).

Let us take a closer look at this greeting case. Prior to encountering another tribe, and seeing that when two persons from that tribe face each other and nod heads, it may never have occurred to us to think of anything that we do as a greeting, let alone a particular mode of greeting. Does that mean that there was not something there that we were used to doing? This is where a theorist may be tempted to say there was, but that we were not aware of it, but to this we need to reply: this is a misleading temptation, for it is one that tries to make it as if the way we ended up thinking of ourselves was what we were always doing. But imagine that when we confront the tribe what we take them as doing is not greeting, but flirting, and thus we come to think of ourselves as flirting by touching and them as flirting by exchanging glances. You might say that what this shows is that there always existed in our group the convention of flirting by touching, but this (as should be clearer now) is a faulty argument. The point is that every time we think something is revealed to have always been so, we forget that things could have been otherwise – that how we may think of ourselves and others is not pre-determined (we just have this peculiar tendency to confirm, to somehow dig in, the way we may currently describe ourselves).

The effect of seeing this should be to see that appealing to convention is – if analysed, as it should be, by reference the contingency of attitudes – of limited explanatory value: there is not something that can be asserted (or assumed) to be a convention (a rule with the status of conventionality in Marmor's language) on the basis that persons would react in a certain way to what is assumed to be an alternative (precisely because how persons will react is not pre-determined, so we cannot rely on a thought experiment as to how we think they are likely to react). Instead, the best way to think about conventions is as contingently useful normative resources, i.e. inherently limited specifications of required behaviours that we may sometimes use to guide ourselves and to evaluate ourselves and others (and not, as Marmor seems to assume they are, repeated behaviours that are asserted or assumed to be conventional

on the counterfactual basis that persons would exercise the attitudes that would confirm the theorist's characterisation of that behaviour as conventional).[170]

Marmor appears to be aware of some of the problems with the assumptions he makes. Thus, for instance, he says with respect to a rule being an alternative that it 1) must be an alternative 'that could have been followed…without a *significant* loss in its function or purpose' (Marmor 2009, 9; emphasis added); and 2) that it has 'to be supported by the same reasons or functions that the original rule serves for the relevant population' (Marmor 2009, 10). Notice that such terms as 'significant' point to the need for evaluation – but what Marmor wants to do is to establish something as an alternative without needing the members of the group to actively and creatively evaluate something as an alternative, and that is why he needs to point to another criterion, in this case 'function or purpose'. However, his very own reference to the term 'significant' (in this case, 'significant loss') suggests that there is no higher level (referred to by Marmor as the level of functions and purposes) that simply determines whether something is an alternative, for one needs to make an evaluation whether that something contributes in the 'same way' or in a slightly different way (and where different, to evaluate whether that difference is 'significant' or relevant). Marmor seems to hide the need for an evaluation by always receding to an allegedly more fundamental criterion that is assumed to determine the status of something, i.e. the status of something as an alternative determined, as it were, by some higher order level of functions and purposes. As he does so, he forgets that at some point there must have been an evaluation made; we say 'forgets' here because he describes matters in such a way that members of a group would just be presented with something that already really is an alternative (the description of the case makes us forget that the members of the group need to see it as an alternative, which also means that they may not see it as one).

B. Constitutive and Deep Conventions

We have now given an outline of Marmor's general definition of conventionality, but what we have seen is that what Marmor has done at every turn is to appear as if he is taking attitudes seriously, but in fact to simply presuppose the nature of something

---

[170] To revert back to the above example: perhaps we would say that hand-shaking is a kind of greeting, but we would say that it is 'disgusting' and that 'head-nodding' is sacred to us. The point, again, is that Marmor assumes something is a convention, and then works back to characterising what he thinks the attitudes must be (to 'confirm', as it were, that his characterisation was correct).

persons do as conventional. Marmor has begun with what he has learnt to treat as a convention (e.g. shaking hands) and he has worked backwards to make conventions out into a much larger class than is warranted if we take attitudes seriously. We have seen that it is only when we fail to see Marmor's explanatory strategy that we think that he has proven to us that there are things that we do that are, by nature, conventional, whether anyone has ever realised it or not. What we have tried to show (in critiquing Marmor) is that describing something as conventional is evidence of a certain attitude to it, and one that is dependent on treating something as an alternative. This means that prior to the forming of the attitude we must have understood ourselves as doing something (e.g. head-nodding as a kind of greeting), which is an understanding of ourselves we may have formed by an encounter with (what perceived, but also may not have perceived) as an alternative way of doing the same thing (the point being that we may not have so reacted).[171] Marmor, however, does not want to make conventionality dependent on contingently exercised attitudes, because this would greatly restrict their explanatory scope.[172]

That Marmor does not want to restrict the explanatory scope of conventions is seen most clearly by his positing the existence of constitutive and deep conventions (which, as we shall see in a moment, are deemed to exist at ever more fundamental, more foundational, levels, which themselves trickle down to determine what counts as conventional, thereby increasingly side-lining attitudes). To see this, let us look first at Marmor's notion of constitutive conventions, and then at his sense of deep conventions.

The constitutive strategy will be familiar to us from our discussion of it in Part II of this thesis (in Chapters 7 and 8), but our task here is to see how Marmor uses the strategy to push our understanding of conventions in a certain direction. The general definition of constitutive rules that Marmor works with is 'those rules that constitute a type of activity, a social practice' (Marmor 2009, 35). Constitutive rules come in two varieties: first, conventional; and second, institutional. Examples of social practices

---

[171] Again, the point here is that there is no underlying nature to 'what' we do: our attitude does not reveal that we were doing something that was conventional, or that it was something more than that, e.g. part of our essence; instead, we exercise a creative ability to understand ourselves in a certain way and an equally creative ability to adopt an attitude to 'what' we understand ourselves as doing (it also follows that not all of us may understand 'what' we do in the same way, or that, even if we do, that we will share the same attitude to it).

[172] It would greatly restrict their explanatory scope because we could no longer speak of an underlying order of behaviours, and an underlying order of reasons for following them (where the attitudes are relegated to being potentially exercised if the need arose).

constituted by conventional rules (and what Marmor sometimes refers to as conventional practices) include: 'structured games, like chess or football, forms of art, like theatre, poetry, or symphonic music, social ceremonies, like a wedding or an initiation ritual and, of course, certain aspects of language use' (Marmor 2009, 31); later, he adds '(some) practices of etiquette (like table manners, greeting conventions)' (Marmor 2009, 35). Social practices constituted by institutional rules (and thus sometimes referred to by Marmor as institutional practices) include 'legal institutions (like legislatures, courts…), quasi-legal (or derivative) institutions (like a college or university, political parties…), and religious institutions (like a church, a congregation)' (Marmor 2009, 35). Having made the distinction, Marmor then focuses on constitutive practices, or social practices constituted by conventional rules, relying particularly heavily on the example of chess.

The rules of chess, says Marmor, are constitutive conventions. 'Their function', he continues, 'is to define what the game is and how to play it: the conventions define what counts as winning and losing a game, what are permissible and impressible moves in the game, and so on' (Marmor 2009, 36). We have already had occasion to reflect, in great detail, about the way of approaching the understanding of games (including chess), so we shall not dwell on it here (see Chapter 7). But what is interesting for our purposes is where Marmor takes this next. Marmor argues that the constitutive conventions also 'at least partly define, or constitute, some of the values we associate with the game and a whole range of evaluative discourse that appropriately applies to it' (Marmor 2009, 36; it is significant that he does not explain what he means by 'partly'). Here is how he elaborates on this in the case of chess:

> To begin with, the rules constitute chess as a competitive game that one can win or lose; and the rules are such that they constitute a particular kind of competition between the players that is, basically, the point or purpose of indulging in the game. It is the kind of competition that values certain intellectual skills of strategic computation, memory, and such. (Marmor 2009, 36)

The constitutive rules of chess are now seen as not only defining the game of chess and how to play it, but also as constituting the values of the game. What might it mean to say that the values of the game are constituted by the rules? The way that Marmor answers this question is to appeal to a contrasting case, but where the

contrasting case depends on a different set of rules: 'imagine, for example, a game like chess that is designed in such a way as to last indefinitely, without the possibility of winning it'; such a game, says Marmor, constituted by different rules, would also establish 'rather different kinds of values than the ones we associate with chess as we know it' (Marmor 2009, 37). What this contrast case allows Marmor to do is avoid addressing the obvious question: is saying that the values of the game are constituted by the rules not saying that one cannot associate any other values with these sets of rules, and if this is what it is saying, is it a plausible thing to say? In other words, when Marmor contrasts two different sets of rules in support of his claim that rules constitute values, he masks the tenuous link between rules and values. We can see this immediately when we consider what other values may be associated with chess, and thus introduce a different sort of contrast case, i.e. not one where there are different rules, but one precisely where the rules are the same but they are experienced differently: for example, consider two friends playing chess in a pub. Imagine that we describe their behaviour in light of the constitutive rules (that define the game and how to play it – so imagine we take for granted this claim about the constitutive rules of chess): does it follow that the two friends must play a competitive game where all that matters is who wins or loses, or is it possible that the two friends desire to have fun in playing an amusing or beautiful game? If the latter is possible, then is it not also possible that the skills important to such an endeavour are not 'strategic computation' or 'memory', but instead, for instance, playing dangerously and with flair (being able to take risks)?

Marmor could respond to this by saying that within the constitutive rules of chess he meant to include the rule that the game is to played competitively, i.e. with the aim of winning (that that is what it means to play chess). Saying this, however, would not work because Marmor could not say that and then also claim that the rules constitute the value of competitiveness, for he would have already defined the rules to include that value.

Another reply might be: but you have pointed (in your example of the two friends in a pub) to a very idiosyncratic use of chess, an exception (so to speak) that proves the rule. But to say this is just to thump the table: it is to insist that the only value that can be associated with the game is the competitive one, and that playing chess in some other way is some kind of anomaly. But notice that even if we accept Marmor's (potential) complaint and focus only on examples of competitive chess, the

claim that the constitutive rules constitute the values of chess is still too strong, e.g. tournament chess often includes two players playing (because of their position in the tournament table, say) to draw (and not to win or lose), and it still may include players who do not care about winning or losing but precisely about playing a beautiful or amusing game.

If, then, Marmor's thesis about the constitutive rules constituting values were to be valid, it would mean we could not enjoy chess in any other way than by competing to win, working hard at home in between competitions on our strategic computation and memory; but clearly we do enjoy chess in other ways, so the claim cannot be valid. Indeed, this result should also make us reflect on the constitutive strategy in general, i.e. if thinking that some 'social practice' (e.g. chess) is constituted by rules also makes one think that when one plays chess one must play it in a certain way (e.g. competitively, in order to win or lose, and relying on strategic computation, etc), then when one shows that this greatly narrows what one can see people who play chess caring about (e.g. playing to draw, or to play a beautiful or amusing game), then this reveals the explanatory poverty or at least limits of appealing to constitutive rules as a way of understanding activities.[173]

Marmor does acknowledge that someone might worry about how he conceives of the relationship between constitutive rules and values; he says, for instance, that someone 'might object that values are just not the kind of things that can be constituted by rules, conventional or other' (Marmor 2009, 38). It is difficult to grasp what objection Marmor has in mind here. Let us try to understand it by looking at Marmor's replies. The replies come in the form of two different readings (one weaker, and the other stronger) of the relation of constitution between the constitutive rules and values associated with the practice.

The weaker sense of constituting value is to 'say that the value of following R under circumstances C is contingent on the fact that R forms part of a system of rules, S, that constitute a (presumably valuable) practice. In other words, following R is valuable only in the context of S' (Marmor 2009, 39). With respect to the chess example, this would play out as follows: 'there is nothing valuable in following the rule that requires moving the bishop diagonally; it is only valuable to follow this rule in the context of playing chess, as a whole game, so to speak' (Marmor 2009, 39). But

---

[173] More positively, it reveals that we need to do more work on the variety and tentativeness of normative attitudes.

this is not an informative reply, for it seems to say only that it is valuable to follow the rule of moving the bishop diagonally when one is playing chess, and only when one is playing chess; this surely is just another way of saying that insofar as one wants to play chess according to the rules, one must follow the rules, including the rule of how the bishop moves.

The stronger reading then says this: 'We could say that a system of rules, S, constitutes a value in the strong sense, if it is the case that engaging in the practice constituted by S is valuable (at least for those who engage in it) in ways in which it *could not have been valuable without the existence of S*' (Marmor 2009, 39; original emphasis). Matters become more complex in this context because Marmor then agues that there are two readings of the stronger reading: an 'ontological' and an 'epistemic' reading. On the epistemic version, 'it is possible to claim that a social practice enables us to realise or appreciate values that we could not have otherwise come to appreciate'; the ontological reading is that 'the practice actually creates values that otherwise could not have existed without it' (Marmor 2009, 39). Marmor does not want to place much reliance on the ontological reading (he says that it is possible), but he does believe in the epistemic one: 'even if it is the case that social practices only enable us to appreciate values that otherwise we could not have come to appreciate, the main thesis about the relation between constitutive conventions and values remains' (Marmor 2009, 40). But both the ontological and the epistemic readings are either false or trivial. Take the appreciation of strategic computation. Clearly, we appreciate strategic computation in many other areas of human endeavour, and it is very plausible that we would have even without chess: one might say, for instance, that strategic computation is involved in solving maths or physics problems, and that these were around a great deal of time before chess ever came to exist. Clearly, an attitude that valued strategic computation could have existed without chess. So the claim is false. But the claim could also be trivial because Marmor could say: we could not normatively appreciate 'strategic computation in chess' without chess (that is indeed how he puts it: Marmor 2009, 39), but this like saying we could not normatively appreciate chess without chess (because he has already defined chess as requiring, by necessity, the use of strategic computation).

Marmor has one more card up his sleeve in this respect (i.e. with respect to the alleged constitutive relation between constitutive rules and values), and he pulls it out in chapter six. This is what he says:

It is certainly possible for various participants in a conventional social practice to have different conceptions of the values the practice is there to instantiate. There is no need to assume that a conventional practice cannot flourish unless all of its participants share a unified conception of its point or values. But, roughly, they must share a certain minimal level of commitment to engage in the practice as it is constituted by its rules, and they must share some plausible conception of the point of the practice and its basic values. (Marmor 2009, 135)

Here, Marmor seems to acknowledge the possibility of some variety in how persons enjoy chess, but nevertheless wants to hold on to some 'minimum level of commitment', some core that must always be there. This is a somewhat softened version of the claim that there are determinants that we must share, for it allows for some diversity, and therefore some contingency in what we might share, but it nevertheless retains a minimum core. Here we must stand our ground: there is no minimum core (or basic set) of values that we must share when playing chess. In fact, it is more plausible to assert the alternative: like many other activities, chess is one that we can engage in and enjoy in an enormous variety of ways, each of us potentially differently but also potentially in a similar way (the point is that there is no need for necessity here). There need be no convergence – though there may be, on any one occasion – as to the 'point of the practice' of chess (i.e. we may, on any one occasion, both treat chess as having a certain point). Of course, there being divergent conceptions of the point may sometimes result in conflict, e.g. one of the friends in the pub wants to play competitively, and the other wants to play a beautiful game: one could get frustrated with the other for taking the game too seriously (or not seriously enough), but then this does happen, and it is important to make room for it happening.[174]

At one point, Marmor puts aside the issue of how constitutive rules constitute values, and instead focuses on how constitutive rules (of chess) are also conventions. He needs to discuss this because it seems counter-intuitive (presumably it is also the reason why Lewis did not) associate the constitutive strategy with conventions, for conventions seem arbitrary in a way that constitutive rules do not. It is in pursuing this problem that Marmor ends up creating the need for 'deep conventions'. This is

---

[174] If it can happen with chess, there is all the more reason to think it happens often in many other contexts of social interaction, e.g. we can debate a philosophical point together and yet have very different conceptions of the point of philosophy.

because what Marmor has to argue is that there is some room for flexibility, i.e. it is possible for the rules of chess to have been otherwise, but still be sufficiently alike, such that there is no 'significant loss in the purpose or function for those rules' (Marmor 2009, 43). Recall that he has to say this because of the way he defined condition 3 of conventionality (the arbitrariness condition). In other words, Marmor has to assert the existence of a higher level – in his case, the level of purposes or functions – according to which something could have the status (independent of our thinking so, for recall that our attitudes are given marginal, *potential* roles) that it is sufficiently alike the rules of chess as we know them to count as an alternative. Proceeding in this way, as Marmor knows, raises a problem for the constitutive strategy, or seemingly so, for the constitutive strategy says that the rules constitute the game of chess (chess could not be played without following the rules). Marmor tries to solve the problem, and escape the difficulty, by appealing to deep conventions.

Marmor invokes deep conventions by saying that 'there is more to the practice' (of chess) 'than just following its rules' (Marmor 2009, 40). This might look like a concession of defeat, i.e. an agreement that the rules do not constitute (or define the game) for there is more to the game than these particular rules, but in fact the way Marmor treats it so to dig the constitutive strategy in deeper: i.e. when Marmor says there is 'more' (to chess than the rules) what he means by 'more' are precisely a set of conventions at a deeper level (the deep conventions, at the level of purposes and functions), according to which the rules of chess are as they are (but according to which they could also be a little different).

What we see here is that the same move is repeated that we saw above when discussing condition 3 (of the general definition): instead of allowing for the contingent exercise of attitudes, as when we might evaluate something as either an alternative or not, Marmor presents things in such a way that he posits a deeper level that determines what would count as an alternative (because the deeper level is imagined as a set of criteria according to which something would be sufficiently alike – i.e. without significant loss in purpose or function – to be an alternative).

In the context of chess, then, the rules of chess – as they currently are – are referred to by Marmor as 'surface conventions', and these are seen to be instantiations of 'deep conventions'. It is important to see the structure of the argument here: as soon as the opportunity arises for introducing some diversity (that we could enjoy the game differently, or that the rules could be different), that diversity is (and those

differences are) said to be but mere instantiations of something, at an allegedly deeper level, that we must share.

So what are the deep conventions that instantiate the surface conventions of chess? Marmor says this:

> Consider…the rules constituting the game of chess. As we have seen in detail, the rules constituting chess are constitutive conventions. They constitute what chess is and how to play the game. In part, they constitute the point of playing the game and some of the specific values associated with it. But all this is possible only against the background of some shared normative scheme about what competitive games are. Such normative backgrounds are what I call deep conventions. Chess, as a game of a particular kind, is only an instantiation of a more general, norm-governed, human activity that we call 'playing a (competitive) game'. (Marmor 2009, 59)

Marmor offers the following list (not meant to be exhaustive, but certainly paradigmatic) of the deep conventions that, for him, 'determine what games in our culture are' (Marmor 2009, 60):

> They are typically rule-governed (even if the rules are rudimentary, tentative or in flux); the rules constitute what counts as winning (and losing), or at least, what counts as success in the game; games concern artificial interactions, with a certain element of detachment from real-life concerns; and they normally involve a certain demarcation of participants distinguished from non-participants. (And there may be other similar features.) (Marmor 2009, 100; see also the list at Marmor 2009, 60)

The idea here, to recall, is that without sharing these deep conventions, we could not play the game of chess (for chess would not be a game – it is only a game because its rules instantiate the deep conventions of what counts as a game in our culture).

One can tackle the plausibility of this idea from two directions: first, one can say that the above argument only works if we agree that chess is somehow necessarily a competitive game (that the rules make it so), but of course we have already argued that the rules do not dictate (even a minimal core) of how the game can be played and enjoyed. Indeed, the very fact that chess could be enjoyed differently suggests that there are no fixed ways in which games must be in our culture. Since we have already argued this, we shall not repeat ourselves again. But there is a second way: it is to say that although chess (understood in a certain way) may offer a plausible example of a game that can be said to instantiate a certain set of (deep) conventions about what

count as games in our culture, there are many other examples of games that show that there is no layer below, no set of conditions of possibility that any game must correspond to in order to be a game. Of course, the difficulty with doing this is that Marmor can try and reply to any counter-example by saying that it is an exception, or not 'really' or 'genuinely' a game, but then it needs to be said that he would be effectively thumping the table, i.e. asserting that only the conditions he identifies are the ones that count.

Consider, then, playing in a playground: here, as two children play they may not be holding still or fixed any particular way that will count for 'success' in the game (let alone any way of winning or losing); imagine the two children are going along amusing themselves, one of them pretends that some rock is some character and speaks on its behalf, and the other responds by, say, talking to the character; now of course, they could continue on like this and then you could say they were playing that particular game where what counted as success was the first person picking some element of the environment and pretending it to be a character and the second person responding to that pretence, but if you did that you might miss that something else might happen, i.e. instead of responding to the character by talking to it, the second person might instead pick another element in the environment and take the lead by creating a character of her own; it is possible here that the first child, if they hold on – try to control – what counts as 'the game' (i.e. the one where they always get to create the character) might complain to the second child: you are not playing 'the game'! But, of course, she also might not, and instead, she might create another character in reply (the point is the first child, in replying this way, is relating with the other in such a way where the other can create 'what' game is being played).[175] The example shows us how two children can play and how they can play without there being some fixed criteria for what counts as success (or winning or losing). It is also shows that when one sticks to there being some fixed criteria one is effectively relating with another person by either projecting that fixation onto how they ought to behave or assuming that both persons must play 'that and only that game' – and, in either of those two senses, precisely not playing with them.

There is yet another way to argue against the idea of deep conventions, and this is inherent in the very idea that they are 'conventions'. For recall that for them to

---

[175] This is not to say that both or either must be relating with the other with a view to constructing 'the game' that they can then conceive of as playing together – they can just be playing.

be conventions, it must be possible for them to be otherwise. Marmor vacillates on this a bit, for he says that unlike surface conventions, deep conventions are 'much more durable and less amenable to change' (Marmor 2009, 59), but change they must be able to for they are (according to his own way of thinking about them) 'conventions.' But now exactly the same move is performed by Marmor that we saw performed when he tackled constitutive conventions: he says that the deep conventions are themselves responses to or they serve 'relatively *basic functions*' (Marmor 2009, 58; emphasis added; we get no insight as to what Marmor means by 'relatively').

One can see the pattern here: first, there are constitutive conventions, but below them are, second, deep conventions, and below these in turn are, at a third layer down, basic functions. In order to avoid the contingency that would result in us either treating or not treating something as an alternative, and thus in order to avoid conventionality being a matter of our attitude to something (where we would be creating the conventionality of something, which would mean that nothing is conventional by nature), Marmor always appeals to a layer below, a set of conditions of possibility or necessarily shared criteria ('normative background' as he puts it), which determines what could / would be counted as an alternative. The same strategy can be repeated over and over again until Marmor would eventually have to confront the basic question: do we end with ever more conventionality (conventionality all the way down, so to speak) or do we end up with some unchangeable foundations, some fixed set of criteria (some 'absolutely' basic functions, etc)? Marmor never answers this, and he cannot, because his entire explanatory scheme depends on repeating the same move, of explaining change as but the inevitable product of something allegedly more basic, of explaining diversity as but the unavoidable result of the conditions of possibility established by some unity.[176]

Perhaps we can give the move that Marmor keeps repeating (of pushing conventionality back one more step, and then one more again…etc) a more positive spin. We can say that instead of reading it as a matter of showing that change is always pre-determined, or diversity is only ever a manifestation of an underlying unity, what his move lets us see is that we can always think that things could be

---

[176] This – in effect – is not to explain change or diversity, but to explain them away; the whole point of our critique here is that change and diversity can only be made room for where room is also made for contingency, for things really being able to be otherwise in ways that are not pre-determined.

otherwise, and that this thought can help us see the limits of any explanatory scheme that we might devise or the limits of our reliance on any particular set of normative resources. Consider how this might work in the context of the following example provided by Marmor. Marmor says at one point that whereas Christian art was figural, and Islamic was not (both of these characterised by him as surface conventions), they both instantiated the deep convention that the visual arts were to function as 'a religious tool of glorifying God and vividly telling the story of the Bible' (Marmor 2009, 61). The positive spin here would ask us not to stop, i.e. to keep the move going, which also allows us to see that what we are calling a deep convention is but a way to understand – to group together – certain things. We would keep it going by, say, increasing the pool of comparison to modern art; as soon as we do so – e.g. as soon as we include within the things to be thought about the urinals of Marcel Duchamp ('Fountain', 1917) – we might say that all three (Christian, Muslim and Modern) are all different ways of, say, expressing our inner frustrations or hopes. The effect of this is to make us see Christian and Muslim paintings in a different way. Of course, this does not mean that we now commit ourselves to saying that the nature of art is to express our inner frustrations or hopes, but just that we treat that as unchangeable for the purpose of grouping things together and thus seeing diversity where we once saw unity, change where we once saw fixity. The example Marmor uses is interesting because it is one where we are confronted with how differently groups previous to us saw the arts; the danger is in thinking that this is really what art must have been for them, for that is where we come to think that they were governed by a shared normative background that meant treating art in some way (i.e. as a glorification of God). But what we are suggesting here is that we can read the example to show not that we have discovered something about how that group really treated art, but rather that we have learnt something about the fragility of our schemes for understanding others (and ourselves); the lesson to draw is that we could be missing a great deal about the paintings of the Christian and Muslim world if we think of them only as glorifications of God (and not, for instance, as expressions of inner frustrations or hopes). The conventionality move, then, is something that reflects more about our practices of describing and explaining, than it proves that we, or some other group we are trying to understand, really are governed by something genuinely and (ultimately, or absolutely) shared (i.e. of trying to explain away change and

diversity by claiming it is but an instantiation of something at a deeper level that is foundational and common).

This completes our discussion of Marmor; there is much more in the book, of course, that we have not discussed, but it is hoped that sufficient detail has been given of his view to show that it does not support the idea (which we should resist) that there are determinants of what is appropriate (including, for instance, what makes participation in certain activities valuable) and what is inappropriate that we must necessarily share. Let us continue now with our critique of this third element of the governance view by turning to the concept of social practices.

Chapter 18. Social Practices

We have already had occasion many times during this thesis to refer to the obstinate presence of the image of games (understood and used, or better abused, in a specific way). The appeal to games can also be encountered often in theories of social interaction, as when such interaction is explained as possible on the basis of, and then described as a matter of following / being guided by / having internalised, a set of rules for interacting. A similar strategy is met with reference to the idea of 'practices', i.e. roughly, of there being distinct (distinct because identified / constituted by distinct sets of rules or norms) domains of interaction where there are established means for achieving established ends. Both of these devices – games and practices – share a certain way of proceeding to account for the normative dimension of social life, i.e. they begin with and construct a rule- or norm- defined structure, and then describe how that structure is absorbed or comes to be internalised in persons who are seen as (on the whole) conforming to the rules or norms of that structure. We need to examine this general way of proceeding here. We will do so first by looking at the idea of practice in this Chapter, and second, in the next Chapter, by considering the appeal to games as ways of understanding / modelling social life. We will, of course, given limited time and space, need to be very selective.

A. The Concept of a Practice

One of the most prominent and sophisticated defences of the notion of practice (building on, but arguably making more plausible, some of the claims made by Brandom) is the work of Joseph Rouse. Rouse has recently (2007a) authored a long, but instructive, summary of his view, and we will need to examine this here. He has also, over the years, engaged in debate with a long-standing critique of the idea of practices, Stephen Turner, and we will see that noting some of the moves made in their recent exchange (Rouse 2007b; Turner 2007a) will allow us to examine some of the principal strategies and stakes at play in this context.

Rouse reminds us that the notion of practice has received extensive attention and application in philosophy and the social sciences, ranging from the Kabyle gift-exchanges discussed by Pierre Bourdieu to the practice of science (in Rouse's own work). Rouse enlists the resources of a great many theoretical contributions to the notion of practice, including Heidegger, Wittgenstein, Foucault, Dreyfus, Giddens,

Oakeshott, MacIntyre, Polanyi and Brandom, thereby broadly grouping together ideas such as language games or ways of life, the background, tacit knowledge, and tradition, amongst others (Rouse 2007a, 639-640). This may seem like too diverse a group to form a coherent theoretical concept, but the important contribution of Rouse's work has been to attempt unification.

According to Rouse, practice theory begins with an emphasis on 'rules, norms, conventions or meanings' (Rouse 2007a, 642). The key idea of practice theory is that 'society or culture is the realm of activities and institutions governed or constituted by rules, of meaningful performances rather than merely physical or biological processes, or of actions according to norms rather than (or as well as) causally determined events' (Rouse 2007a, 642). It is because of the emphasis on rules, norms and conventions, and the resulting picture of social life as governed by them, that Rouse's reconstruction of practice theory is of obvious importance for this thesis (as an instance of the governance view).

As Rouse presents it, the central problem of practice theory is how to present persons as guided by rules, norms or conventions but without either positing an infinite regress (of rules for interpreting rules, etc) or introducing an implausible picture of the mind (where implausibility is seen to consist in a picture that is overly deliberative and self-conscious). Hence, for example, the reference to and interpretation of Wittgenstein as a figure who showed how there was a way of following rules blindly (i.e. without interpretation) – a way of 'grasping rules without interpreting them, which is "exhibited in actual cases"' (Rouse 2007a, 642). Hence, too, the reference to Heidegger, who is enlisted for providing a picture of 'understanding' as a more basic notion than interpretation, where the latter is described and explained as 'only possible against the background of a prior understanding of the situation' (Rouse 2007a, 643). Rouse here recalls Heidegger's famous example of hammering, and summarises Heidegger's reading as follows:

> In Heidegger's example of hammering, one must already understand the general context of carpentry (the relation between hammers, boards, nails, buildings or furniture, and the various purposes they serve), one must have a sense of how to proceed (hammers must be picked up to be used, held by the handle, swung rather than thrown, hit the nail on the head rather than the shaft and so forth), and one's interpretation is governed by a general sense of what would bring it to fulfilment or completion. Without some prior practical grasp of these considerations, nothing one does with a hammer could amount to

hammering with it (indeed, there could be no hammers without such an understanding of hammering). (Rouse 2007a, 643).

We have, throughout this thesis, been resisting such a picture, for, as we have seen, it tends to picture the mind learning and working in too passive a way, positioning it as facing an environment that is said to contain within it certain possibilities for action, with the effect that the mind is pictured as always and already geared towards detecting and responding to those possibilities (and enacting them out). We have been examining how structuring the environment on the basis of there being such possibilities (often explained as either embedded / implicit rules, norms or conventions, or as abilities / dispositions / sensitivities to following them) means that we miss the sense in which the mind may create possibilities as it proceeds, and thus how it is misleading to position the mind in the context of an environment always and already organised for the mind in some way.[177] We have seen how adding the normative dimension can all too easily result in the elimination of the ability of the mind to create possibilities for itself,[178] for adding the normative dimension often results in arguing that insofar as the mind is capable of acting appropriately (and being evaluated as having made a mistake), it must be describable and described as being guided (even when it does not realise it, though 'realises' at least potentially as reflected in reports / justifications) by some rules, norms or conventions (i.e. determinants of what is appropriate and what is inappropriate).

This is how Rouse articulates the general strategy of practice theory: 'The concept of a "practice" is widely invoked in social theory to identify the locus of this background understanding or competence that makes it possible to follow rules, obey norms, and articulate and grasp meanings' (Rouse 2007a, 643). What we have been saying in reply to this is that 1) one need not, and ought not, to describe the environment in which we act as already and exhaustively structured by rules, norms and meanings; and 2) that when one does so, the effect is that the mind and its abilities are pictured in too passive and rigid a fashion, i.e. where the mind is pictured as learning something specific that it merely recognises and re-enacts when faced

---

[177] This can also be re-phrased in terms of person-to-person relations, in which (from the second-person perspective) the other person can create unfamiliar (to us) invitations or opportunities to relate (these invitations or opportunities being precisely possibilities for interaction created in the process of relating with an other).

[178] Or, once again, in a person-to-person context, of eliminating the ability of one mind to be responsive to the possibilities created for it by another.

with (what is assumed / presented to be) the relevant environment in which that something is there to be recognised and re-enacted. In effect, by appealing to 'background understandings and competences' prior to, and thus enabling, the following of rules or norms, practice theory presupposes what we have been arguing cannot be presupposed.

B. Challenges for Practice Theory

Rouse's paper, however, deserves closer analysis because Rouse is alive to the difficulties with the notion of practice. Addressing these difficulties assists Rouse in providing a more flexible notion of practice than is often found in the literature. We need to discuss how Rouse presents and attempts to solve the difficulties, and whether this is sufficient to make the notion of practice plausible. The following are the difficulties discussed by Rouse: first, the sense in which practice theory attempts to move beyond the split between agency and structure in social science; second, the sense in which practice theory deals with issues concerning the transmission of social knowledge; third, the sense in which practice theory takes the body and embodiment seriously, and thus is not an over-intellectualised account; fourth, the issue of whether, and if so to what degree, the form 'practice' takes is reliant on language; and fifth, the exact sense in which practice theory has its own distinctive notion of the normative. Rouse also covers other difficulties, but these are not as central as those above to our discussion. As we shall see, the key issue will be the fifth – the conception of the normative – and here we shall see that Rouse comes close to articulating a view (though in abstract terms, with few examples) of a sense of the normative that need not result in picturing the mind as passively confronting laid out environments where what is appropriate is to be merely detected and enacted out (indeed, as we shall see, Rouse's discussion of the normative is somewhat at odds with how he defines practice theory in general – indeed, his discussion is one that allows us to show where practice theory is inadequate).

According to Rouse, practice theory sits in between agency and structure because although it asserts that 'practices are composed of individual performances', it nevertheless also argues that these performances 'take place, and are only intelligible, against the more or less stable background of other performances' (Rouse 2007a, 645). The key idea is that backgrounds, unlike structures, cannot be conceived in isolation from 'ongoing performative reproduction', and performances, unlike

agents, cannot be 'properly characterised or understood apart from their belonging to or participation within a practice sustained over time' (Rouse 2007a, 645). At one point, in a further attempt to distinguish backgrounds from structures, Rouse says that unlike structures, backgrounds are dynamic, historical, dissonant, not monolithic, and require 'ongoing contestation and struggle' (Rouse 2007a, 646). In the abstract, these features of backgrounds are promising, but Rouse does not give us any sense of how conceiving of backgrounds in this way changes the way the notion of a practice would describe the normative dimension of social life. This makes it very difficult to assess the precise sense in which change and diversity, including contestation and struggle, are made room for. The task is made doubly difficult by the fact that the emphasis is placed on – as we saw above – the 'ongoing reproduction' of performances. Thus, even though Rouse does make reference (using Sherry Ortner's phrase from her influential paper, 1984) that society is 'made and unmade through human action and interaction' (Rouse 2007a, 646), we receive no further elaboration on the precise way in which it is 'made and unmade', and instead the focus turns to how certain backgrounds are and remain stable (Rouse 2007a, 646-7) and how this stability is transmitted from one generation to the next. What we would need to build on, should we take something from Rouse here, is the potential for a research program that focuses on the way in which 'making and un-making' may not be conceivable along the lines of 'reproduction'. Thus, in discussing this difficulty, Rouse may have succeeded in moving away from the dichotomy of agency and structure (though an advocate of this dichotomy could say that he still reproduces aspects of it by appealing to two different terms, performances and backgrounds), but the manner in which he does so – i.e. by emphasis on reproduction – means that we confront the familiar problem of not making sufficient room for change and diversity.

This first difficulty is connected to the second, namely the question of transmission. This has been a problem for practice theory for a while, and was perhaps most forcefully articulated by Stephen Turner's 1994 attack on the idea of social practices. Rouse argues that there have traditionally been two ways of trying to meet Turner's challenge: one is to appeal to causal connections, and the other to appeal to backgrounds that incorporate understanding. Rouse argues that practice theory at its best offers a hybrid (again, goes beyond the dichotomy) of these two approaches. In other words, practice theory neither settles for the long-standing 'tradition of understanding socialisation into shared practices as a matter of sheer

imitation, training, and sanctions' (which would reduce transmission to 'straightforwardly causal means'; Rouse 2007a, 648), and nor does it present transmission as a matter of 'appropriate uptake, which involves some understanding of the performance to which one responds' (Rouse 2007a, 649). Rather, as for instance in Davidson and Sellars, 'practice theorists…treat what the latter calls the space of causes and the space of reasons as parallel, non-intersecting domains of understanding, such that a theory of social practice could only avail itself of conceptual resources internal to the space of reasons' (Rouse 2007a, 651).

Again, as with the attempt to go beyond the split between agency and structure by using two other labels (performances and background), so here we could raise a doubt as to whether the space of causes and the space of reasons is not just another version of the dichotomy between the appeal to causal connections and the appeal to understanding – in other words, we could make the point that positing a concept (practice) that is designed to make the two interdependent or 'parallel' does not necessarily free us of the dichotomy (for we are still conceiving of the unity as an amalgamation of two different things). But let us put this problem aside, for another one arises. This we can see when Rouse discusses the example of learning a language as a way of showing how practice theory would deal with transmission (in a way that allegedly goes beyond the dichotomy):

> One could not learn to speak a natural language without the capacity to differentiate linguistic signs (phonemes, letters, gestures or whatever serves as the relevant tokens), and the ability and disposition to reproduce them by imitation. Babies babbling and imitating the sounds made by others are not yet language speakers, however. Language is holistic, in the sense that a speaker cannot have the ability to understand and produce one sentence unless she can understand and produce many of them, in appropriately interconnected ways. So having acquired the causally-generated ability to imitate meaningful utterances, our proto-speaker must then somehow be able to pick up on their semantic significance. The realisation of this capacity would undoubtedly require appropriate responses from others (additional utterances to imitate and respond to, but also appropriate corrections of and constructive responses to one's own performances). Speakers characteristically respond to language learners by treating them as if they had a capacity they manifestly do not yet have, by responding to their imitative utterances as if they were already meaningful performances. Yet such efforts to initiate others into the practice would not work unless these cues prompted (rather than merely causally provoking) the right kinds of response from the learner. Training and proto-interpretive or expressive uptake are both necessary, so as to produce not

merely de facto conformity to social norms, but a self-policing conformism. (Rouse 2007a, 650-11)

It is, once again, not clear how this goes beyond the dichotomy, for it still relies on two different modes of transmission (just this time, pictures them as working together), namely imitation (which is causal, and needed at least initially) and 'proto-interpretive or expressive uptake' (which is said to the 'understanding' side, and allow for more complex learning). But the bigger problem is identified at the end, namely, where although the emphasis is not on 'mere conformity', it is still conformity that is the aim, though of an allegedly more complex self-policing kind.[179] This does help defend against claims of autonomous structures somehow existing independently of performances by persons, but it nevertheless pictures persons as but means for the reproduction of linguistic patterns. What this makes difficult to see is, once again, the sense in which we do not, in order to speak a language, need to reproduce the same patterns, but may roughly overlap in our linguistic abilities (the diversity point) and can introduce possibilities for the language that cannot be accounted for as but reproductions of pre-existing patterns (the change point).

The third difficulty for practice theory discussed by Rouse is the role of the body and embodiment. One of the strongest motivations for practice theory is to avoid the over-intellectualised picture of the person who is imagined as constantly deliberating, interpreting and inferring. Again, Rouse here suggests that practice theory at its best does not resolve to think of the body as but a means directed by a mind – again succumbing to another dichotomy (the mind-body dichotomy) – as would be the case were practice theory to appeal to the body as the locus of 'mere dispositions or habits' (Rouse 2007a, 652), but appeals to 'skills' and thus pictures the body as a locus of 'practical knowledge' (as in the work of Polanyi and now most prominently in the work of Dreyfus). Rouse identifies four features of Dreyfus's notion of skills: first, there is the 'implicit "I can" that is the bodily-intentional analogue to the Kantian "I think: that tacitly accompanies all mental representation', the idea here being that any bodily performance 'takes place against a background of a balanced, poised, directed bodily set that enables…effective focus' (Rouse 2007a,

---

[179] The roots of this problem in fact arise earlier in the above-quoted passage, in the conception of language as system of rules governing tokens: the problem, as we saw with Chomsky and Pinker, is that the only change that is conceived is always and already explicable as an instantiation of a pre-existing rule, or if not rule, then principle – one always ends up receding to an allegedly ever-more fundamental level (see Sections 8.C. and 11.A. above).

653); second, the notion of a bodily intentionality that is not mediated by intermediaries (e.g. 'meanings or spatial representations'), which Rouse also refers to as 'skilful responsiveness', and which means that one avoids the need for explicit representationalist accounts for the way that human agents are oriented to objects (Rouse 2007a, 653); third, the skilful responsiveness in question is not a matter of merely repeating the same movements, but something more flexible, which is understood to mean that 'skilled performances manifest a common embodied sense, a directedness toward a goal through varying means' (Rouse 2007a, 654); and fourth, the idea that although bodily skilled responsiveness is important, it is not what we always use, for it is possible, especially when we are novices, that we will engage in 'explicit rule following and merely habitual motions' (Rouse 2007a, 654).

We have discussed the problems here previously (e.g. in our brief discussion of Dreyfus's criticisms of and replies to McDowell; see Section 12.B.), including that the environment is still depicted as laid out or organised in some way, such that the body is seen precisely to be responsive to what is already there for it to respond to.[180] The third feature is not one we have discussed, and it does try to make room for some flexibility, but what is significant is that that flexibility is only depicted in terms of varying the means (where the change itself is but a variation of something pre-existing) towards the same end, as when (in Rouse's own example) 'Having learned to spike a volleyball, I do not do the same thing again and again, but am instead capable of doing something slightly different each time, in response to slightly different circumstances' (Rouse 2007a, 654). On this view, we are still faced with circumstances that have a certain structure – certain ends and certain means for achieving them – with the proviso that we have some room to manoeuvre within the pre-established means. This helps alleviate the passivity of the mind a little, but still pictures it as incapable of creating its own possibilities for action (including its own ends), and the reason why it does so is that it still positions minds (used broadly here to include bodily skills) in a context described by the theorist as already specified / organised in some way.[181]

---

[180] The problem here is not with saying that there are things in the environment to respond to, but with presuming, in describing the case, what those things are: the point is to get as close as possible to depicting the agent relating with the environment (to get closer to this relation, one has to leave room for indeterminacy, in one's description of the case, as to 'what' the agent will relate with in the environment).

[181] We see this passivity at play also in how Rouse discusses freedom as the realisation of possibilities assumed / posited by the theorist to be available to the agent (see Rouse 2007a, 654-5).

The fourth difficulty addressed by Rouse is whether, and if so the degree to which, practices depend on language. Here, Rouse's reply moves through a series of stages. The first stage is the presentation of views that picture elements in the background as 'pre-suppositions'. Practice theorists do not picture these as necessarily linguistically constituted (as say, 'explicitly articulated and accepted propositions or rules': Rouse 2007a, 655) – indeed, they cannot, for avoiding that view is one of the foundational motivations for practice theory – and so they picture them as (at least capable of being) 'unarticulated' and not the product of any explicit agreement. According to Rouse, there are two versions of these theories that appeal to pre-suppositions: first, there is a version that sees these pre-suppositions as playing a justificatory role where they are understood to be 'shared commitments that function[] as justificatory bedrock' (Rouse 2007a, 656); and second, there is the version in which they are understood more in terms of intelligibility, i.e. for instance in the treatment of Heidegger, Gadamer, and Dilthey, where they are understood as 'obvious', 'tacit' ideas, that were so 'embedded in…a social practice that they do not need to be said' (Rouse 2007a, 656-7). On this second version, 'once these presupposition have been brought to explicit attention, their role can shift toward justification, but prior to that the unarticulated presuppositions exists as a limit of intelligibility (Rouse 2007a, 657) – in this sense, the second version incorporates the first, but broadens out the notion of pre-supposition.

Rouse suggests that these 'pre-supposition' theories have problems, which are associated with the fact that 'the notion of "presupposition" suggests some form of semantic content' (Rouse 2007a, 659), and he argues that 'those practice theories who emphasise the bodily basis of practices…often emphasise a very different relation between language and social practice' (Rouse 2007a, 659). This alternative is exemplified, for instance, by Marcel Mauss's notion of ways of walking, or Clifford Geertz's 'cultural exemplary ways of running, squatting, stroking the feathers of a cock, and avoiding bodily acknowledgement with of others' (Rouse 2007a, 659; see Mauss 1979 and Geertz 1973). This skilful, 'practical bodily know-how' 'bypasses any verbal expression, even (or perhaps especially) in the process of its acquisition or transmission…leaving rules behind in order to achieve a distinctively bodily capacity' (Rouse 2007a, 660). Here, we return back to the third point above, namely the reference to bodily skills as in Dreyfus's account – the point being, in this context, that Rouse suggests that rather than try to explain practice via language or linguistic-

like entities, we do better explaining language via bodily capacities. Insofar as we do return to that conception of bodily skills, the same problems we identified above apply here.

But Rouse, in fact, goes further in this context. This may because it is reliance on language (or analogous concepts, such as the notion of a presupposition), that often has the effect that change and novelty is explained away as nothing more than a variation or instantiation of something prior and pre-existing (whether a certain set of fixed meanings or minimum shared contents, or some set of rules or principles). Interestingly, as soon as we start thinking of language as a practice, as Rouse does here, we can no longer appeal to those language-like explanatory instruments, which may mean we can begin to take change, novelty and diversity more seriously. It is in this context, for instance, that Rouse quotes Davidson as saying that 'there is no such thing as a language…a clearly defined structure which language-users acquire and then apply to cases' (Davidson quoted in Rouse 2007a, 662), himself adding 'there is only the activity of interpretation itself…which always outruns any systematic structure acquired or presupposed in advance' (Rouse 2007a, 662). Rouse does not, however, elaborate on this, and nor does he offer examples of how such a view might give us a conception of bodily skills where learning is un-systematic and therefore leaves open the ways in which persons can create possibilities for themselves. Nevertheless, even if only with these hints, Rouse does offer some resources here that we may be able to build on in order to make room for the second-person perspective.

Rouse's discussion of the fifth difficulty contains yet more promising (from our perspective) resources, these also being ones that arguably distinguish him from mainstream practice theory. As flagged above, the fifth difficulty relates to the conception of the normative. At first, Rouse articulates the appropriate position (for practice theory) as one in between what Brandom (1994) identifies as regulism and regularism.[182] And, his position does bear resemblance to Brandom's, but Rouse also says some things that are distinctively his own.

---

[182] Briefly, both regulism and regularism are conceived of by Brandom as responses to the challenge of distinguishing between proper and improper performance. Regulism does so by appealing to 'rules all the way down', and thus including meta-rules that govern how ground-level rules are to be applied. Regulism does not work because it leads to an infinite regress (of rules). Regularism may also refer to rules, but it conceives of these rules in a purely regular fashion, i.e. as governing the behaviour of persons without those persons being able to express them. The problem with regularism is that it has no way of picking out which of the regularities matter for determining what is proper and what is improper about a performance (this is the 'gerrymandering objection' against regularism).

The first step is to say that in between regulism and regularism lies what Rouse calls 'a complex network of mutual interaction' (Rouse 2007a, 671). The idea here, he says, is to move away from a view of performances as 'integrated within the practice…by a shared semantic content or behavioural similarity' (Rouse 2007a, 671), and to focus instead on responses in the course of interaction, such as where one responds to another by 'correcting it, rewarding or punishing its performing, drawing inferences from it, translating it, imitating it…circumventing its effects, and so on' (Rouse 2007a, 670; the list is narrow, but perhaps we could extend it to, for instance, acknowledging another's performance as setting a new standard). This complex network of mutual interaction is not yet 'normative', and so a second step is required; this step asserts 'that these patterns of interaction must constitute something at issue and at stake in their outcome' (Rouse 2007a, 672).

Rouse describes this step in two different ways, and the difference here is crucial. On the one hand – and this is the sense we can build on – Rouse argues that 'what is at stake…is the difference it would make to resolve that issue one way rather than another' (here, when he refers to 'that issue' he is referring to an example in which what is at stake is what it is to be a Jew: see Rouse 2007a, 671-2). Rouse leaves open how something can be resolved, including noticing that there may be alternatives that cannot be pre-described as options available to those interacting. 'Practices', he says, 'point ahead of themselves toward something essentially contestable' (Rouse 2007a, 672). On the other hand, however, Rouse also speaks of what is at stake as being 'constitutive' of practices; as above, he says 'these patterns of interaction must *constitute* something at issue and at stake in their outcome' (emphasis added). It is not clear what he means by this; on one view of it, what he could be read as saying is that any interaction, insofar as it has a normative dimension, must be describable as something that has something specific at stake (e.g. what it means to be a Jew), but with the proviso that any way in which this is resolved is left open. If this is what he is saying, then it is surely too strong: there may be many things – which we cannot describe in advance – that may be at stake in an interaction; it need not be the case that any one interaction is describable as being constituted by something specific at stake. But perhaps what he could be read as saying is that it is only in the course of interaction (and not something imposable from above) that something arises as something at issue and at stake. This reading is supported by the

following passage, in which Rouse describes the third step in the articulation of practices:

> Most philosophical conceptions of normativity presume that there must be some determinate norms that already govern the performances accountable to them, and thus that already settle what is at stake in the practices to which those performances belong. Such conceptions might allow for epistemic uncertainty about these norms on the part of the practitioners, but not metaphysical indeterminacy. Normative practice theories, however, take the issues and stakes in practices to be indeterminate (or perspectively variant), and this amounts to a third crucial feature of their conception of practices. (Rouse 2007a, 672)

Here, Rouse appears to say explicitly that what is at stake cannot be articulated in advance, and, crucially, he extends this limitation not only to epistemic uncertainty, but also to metaphysical indeterminacy. This suggests that it is not only a resolution of what is at issue and at stake that is left open, but also what could be at issue and at stake is as well. This passage is important from the perspective of this thesis. When Rouse adds that this commitment to metaphysical indeterminacy includes a rejection of any '"sovereign" standpoint "above the fray" [quoting Brandom] from which competing political or epistemic claims can be definitively assessed' (Rouse 2007a, 673), he offers further resources for this thesis. The problem, however, is that Rouse does not offer any examples of this way of understanding the normative dimension (of practices).[183]

Further, his general articulation gives us reason to think that at least insofar as he conceives of the normative dimension in the way above, Rouse is not really a practice theorist at all. Recall that the basic formulation of practice theory, at least at the social level, was that it asserted that 'society or culture is the realm of activities and institutions governed or constituted by rules, of meaningful performances rather than merely physical or biological processes, or of actions according to norms rather than (or as well as) causally determined events' (Rouse 2007a, 642). In the above three steps, which outline the normative dimension of practices, Rouse has not once referred to rules, norms or conventions, but rather to dynamics of interaction where what is at stake and how it is resolved is 'always prospective' (Rouse 2007a, 673). It

---

[183] This is not meant, of course, as a criticism of Rouse's oeuvre, for we are limiting ourselves here to his most recent statement of his view. The problem is that it is this latest statement that we find most congenial, but it is also this statement that is low on examples.

does not look at all as if this sense of unpredictable prospectivity, as well as the claim of metaphysical indeterminacy, is at all compatible with the idea that social life is comprised of activities that are constituted and governed by rules, norms or conventions that allegedly rescue the meaningfulness of actions and interactions.

Rouse, however, is difficult to enlist as a potential ally in criticism of the governance view because some ways in which he articulates his conception of the normativity of practices renders it more readable as an instance of the governance view. For instance, when he says 'Normativity is an interaction orientation toward a future encompassing present circumstances *within its past'* (Rouse 2007a, 673; emphasis added), this suggests something very close to the idea that what is at stake and the manner in which it is resolved in the present is conditioned by possibilities existing prior to (even if not known by the participants in) the interaction, which would contradict the above statements concerning prospectivity. Similarly, when he says that 'any effort to stand outside of an ongoing practice and definitively identify the norms that govern its performances is instead incorporated within the practice, as one more contribution to shaping what the practice will become' (Rouse 2007a, 673), it becomes difficult to understand what he means by 'the practice': it seems to turn the idea of practice into something all-encompassing and yet with a boundary. The point, with respect to this last phrase, is: why posit something called '*the* practice' here? What work is this concept doing? Can't we do without it? Certainly, the idea that theories themselves are but attempts – necessarily limited attempts (for more on this, see Rouse 2007a, 663-6) – that may contribute to what we think is at stake and how it can be resolved, is a very promising idea – but why envelop all this into one, unified, specifiable thing, as '*the* practice' suggests it might be? What is the basis upon which Rouse can assert that all these theoretical attempts are themselves 'situated within the field of ongoing activity to which contributes' (Rouse 2007a, 665) – what is '*the* field'? To revert to these devices appears to bring us back to situating us (theorists and participants alike) within some metaphysically established setting; the point being made here is that this neither helpful nor necessary.

C. Turner vs. Rouse

Some of the difficulties and issues we have encountered above are also played out in the above-mentioned debate between Rouse and Turner. The debate is important for our discussion because it focuses specifically on normativity (in this case, on the

specific kind of normativity embraced by Rouse's version of practice theory). As we have seen above, Rouse argues that his conception of the normativity in practices falls in between the regulist and regularist conceptions (that Brandom (1994) also critiques) – the point in this context being that it was, according to Rouse, the regulist and regularist versions that Turner critiqued in his 1994 book.

A large portion of Rouse's short (2007b) paper repeats what we have presented above as the three steps of understanding the allegedly distinct normativity of practices. However, Rouse also adds several paragraphs that further elaborate on his view. The same tensions we saw above are also present in these additions. Thus, for instance, towards the end of his paper Rouse argues there are two advantages to 'conceiving of normativity in terms of how we hold one another accountable to what is at issue and at stake in ongoing practices' (Rouse 2007b, 54). The first advantage is that 'it avoids a widespread and misleading theoretical temptation to posit underlying unity as a condition of understanding' (Rouse 2007b, 54). Rouse argues that this temptation is present both in philosophy and in the social sciences: philosophers of normativity, as well as 'traditional conceptions of culture in anthropology or of structures and norms in sociology' all 'posit stable meanings, rules, norms, patterns, or presuppositions underlying the manifest diversity of social life…[they] identify understanding with repetition of the Sacred Image of the Same' (Rouse 2007b, 54). The second advantage is that it expresses modesty about what philosophers and social theorists can have access to: 'There is', says Rouse, 'no special domain of rules, norms, values, transcendental conditions, natural law, or any other grounds to which philosophers or anyone else has distinctive access…' (Rouse 2007b, 55).

These are very promising ideas from the perspective of this thesis. We have seen how the idea of underlying unity has often led to the description of cases where persons are passively faced with already available choices that they can only repeat / re-enact, and we have also witnessed the postulation of higher order levels of the normative or super norms that ordinary persons do not have access to but that theorists can make explicit.

However, the above additions are also difficult to evaluate, partly because there are other aspects in the added paragraphs that are in tension with them. Thus, for instance, although Rouse says that 'There is no god's eye view that offers a definitive standpoint form which to discern what those stakes really are', he adds that 'There are only ongoing efforts to forge a viable future together from within a shared but

contested past and present' (Rouse 2007b, 55) – and what is difficult here is to see what he means by 'shared', and why he thinks something 'shared' is necessary. Further, although Rouse refers to 'ongoing patterns' he modifies this by positing 'shared' or 'partially shared circumstances', saying further that there are 'pervasive and salient features of the world in which we find ourselves', such as 'languages and social practices', and it is to these that our bodies become 'responsive' (Rouse 2007b, 52-3). Perhaps the following is most problematic phrase: 'Our participation in those practices enables us to become the agents that we are through our mutual accountability to the possibilities those practices make available and to what is thereby at stake for us in how we respond to those possibilities' (Rouse 2007b, 53). Putting matters this way suggests that there are practices, and that there are things at stake in those practices, and that persons become agents by being held mutually accountable to being correctly responsive to those things (i.e. identifying what is there, in those shared circumstances, to be identified). Situating persons in shared circumstances where practices constitute what is at stake certainly reads like metaphysical determinacy. But, again, Rouse is notoriously difficult to pin down, sometimes even within the same sentence; thus, for instance, he says: 'performances are normative when they are directed toward one another as mutually accountable to common stakes, albeit stakes whose correct formulation is always at issue within the practice' (Rouse 2007b, 53) – what makes this difficult to assess is in what sense there can be a 'common stake', and in what sense there can be a 'correct formulation' of it, if what is at stake, and how it can be formulated, is 'always at issue'; it is not clear, once again, what Rouse means by all this being contestable 'within the practice'.

Turner, in his 2007 reply to Rouse (Turner 2007a), does not pick up on these ambivalences in Rouse's treatment of normativity. This is largely because Turner has his own conception of the normative, which treats it as a transcendental claim about why behaviour cannot be reduced to causal explanations.[184] Turner's complaint against Rouse is that he has no ground upon which to claim that normative explanations of social behaviour are 'exclusively valid' (Turner 2007a, 70); indeed, it is stronger: it is that we can do without normative explanations for they add nothing

---

[184] Turner does think that normative explanations can play a part, but only ever as 'false' theories that persons treat as 'true': see Turner 2010 for details.

that non-normative – causal – explanations cannot already do.[185] For Turner, references to something normative contribute 'nothing to the ordinary stream of explanation' (Turner 2007a, 69):

> If we 'non-normatively' say that 'practices' is a term for describing the pattern of interaction produced by the attitudes, beliefs, and habits of the participants, we have described the same thing and done so without leaving any facts out other than the supposed fact of normativity. In short, the normative language of competencies, accountability, and the like, which seems to demand a novel, validity-affirming explication, is merely a way of describing the natural in a normative way, not a novel set of facts and not a novel explanation of any of those facts. (Turner 2007a, 70)

The above is a very strong anti-normativist position.[186] Throughout this thesis we have tried to avoid putting matters in such an all-or-nothing manner, i.e. either with or without normativity. In that sense, we can agree with Turner that insofar as one argues for the exclusive validity of – or the necessity for – an explanation of conduct as governed by rules, norms or conventions (imagined as determinants of what is appropriate or inappropriate), this should be resisted, but this does not mean we need to go so far as to argue that the concept of the normative, and normative resources, play no roles – let alone important roles – both in explanations of social life and in the experiences of living that life. Normative language does add something, just like causal language does – neither is more fundamental than the other, and both can be appealed two in ways in which we make sense of ourselves and each other; the challenge is not to over-estimate and over-generalise the role that either can play.[187]

We have not covered here the range of practice theory views – we had an indication of how broadly this can be construed when we referred to the theoretical influences identified by Rouse – but we hope to have given sufficient indication of the basic themes of this approach.[188]

---

[185] Turner also argues that some of Rouse's language utilises, without acknowledging, causal terms, such as 'maintain' and 'interact' as when Rouse says: practices are 'maintained by interactions' (see Turner 2007a, 69).

[186] In Turner 2010, chapter 6, discussing Davidson and Weber, offers a less strongly anti-normativist account. For a discussion, see Del Mar 2010. For more of Turner on Davidson, see Turner forthcoming.

[187] Again, for details of this criticism of Turner 2010 see Del Mar 2010.

[188] For more of Turner's view on practices, see not only Turner 1994, but also his retrospective look in Turner 2007b; an important historical figure here, especially in terms of sociological theories of practices, is Durkheim: for an analysis, see Rawls 2004.

Chapter 18. Social Games

The notion of practices, which we examined in the previous Chapter, is closely aligned with the notion of games; and both have been applied in similar, though nevertheless distinct ways, to the study of social life. It is time that we now looked at the appeal to games in a social context in a little more detail. We will do so by reference to aspects of the work of Erving Goffman. Before we turn to his account though, it will be useful to consider very briefly the appeal to games – and game-like explanatory devices – to capture the normative dimension of social life. Towards the end of this Chapter we will note some echoes of Goffman in contemporary work, as well as comment on so-called 'breaching experiments', the theoretical background to which has some affinity with Goffman's views.

A. Games and Goffman

It was not to an unenthusiastic audience that Clifford Geertz once declared: 'what the lever did for physics, the chess move promises to do for sociology' (Geertz 1983, 22; quoted in Manning 1992, 58). Indeed, as Benjamin DiCicco-Bloom and David R. Gibson recently put it (2010), 'sociologists study games on occasion, but they liken social life to games with abandon' (DiCicco-Bloom and Gibson 2010, 248). According to DiCicco-Bloom and Gibson, this 'abandon' is explicable on the basis that there are certain properties often associated with games that make the analogy a useful one to appeal to (the assumption being, of course, that those properties are also ones that one ought to ascribe to social life):

> These properties include: that, like games, some areas of social life are rule-governed…goal-oriented…and possibly competitive in that there is some coveted regard that not everyone can possess…; that action involves discrete moves…and perhaps something approximating alternating turns…that skilled 'play' involves the use of strategies…; that with extended play one develops an instinctive 'feel for the game'…; that earlier actions constraint subsequent options for oneself and others…; that a setting may demand that participants ignore external distractions…; that play is engrossing…; that roles are differentiated and knowingly co-ordinated…; that rules are arbitrary, in the sense that they are based on convention rather than physical laws…; and that, as a consequence of rules' arbitrariness, many different games are possible. (DiCicco-Bloom and Gibson 2010, 248)

DiCicco-Bloom and Gibson cite a wide range of authors correlated with the above properties, including Bourdieu, Goffman, Elias, and Mead. Their own project is to argue that the trouble with the appeal to the games analogy is that the analogy is often made to 'any particular game' where what theorists appear 'to have in mind [is] some sort of prototypical, Platonic game' (DiCicco-Bloom and Gibson 2010, 248), and they go on to point to differences between the games of chess, go and Texas hold 'em poker. DiCicco-Bloom and Gibson's point is a good one, and they are surely right that more work needs to be done on the differences between games, but we should also add that it would be exaggerating matters to assert that all social theorists who appeal to games are all committed to the above properties of games. Perhaps the one overall strategy common to the appeal to the analogy is that games offer a seemingly obvious example of a shared stage, or shared environment, on the basis of which persons can be described as interacting in a certain way. They offer such a shared stage or shared environment because they appear – at least from a distance that does not question it – to be fairly autonomous, isolated domains structured (defined, made possible by, constituted by – depending on the terms one uses) some set of rules (or norms, or conventions, or sometimes a bundle of these) that enable certain actions to be repeated over and over (and yet with variety, i.e. as instantiations of the rules, much like the rules of chess provide almost infinite possibilities for playing the game). Games so understood thus appear to offer examples of clean-cut explanatory devices. In this general sense, then, although there is room to resist some aspects of the analogy – for instance, by saying that, as DiCicco-Bloom and Gibson assert Garfinkel did, 'the rules of social life are rarely as clear as those of a game' (DiCicco-Bloom and Gibson 2010, 248) – the overall idea, i.e. the idea that participants in social life engage in rule-governed contexts (where the theorist may need to tell us what the rules are, even if they are unclear to the participants), remains intact. This is also why the appeal to games makes it an instance of beginning with – presupposing or postulating – a social order within which persons are imagined as participating (including learning to participate in until they become social experts, i.e. masters of the rules of social life).

It would take us far beyond the limits of this thesis to provide a history of the appeal to games in accounts of the nature of social life, including its normative dimension (though we will be briefly returning to the contrast between games and play, and thus the distinctiveness of playing, in Part V of this thesis; see Chapter 20).

But given our focus on Goffman, it will be useful to give some more background, especially where this background is of relevance to understanding Goffman's work, as this has been helpfully described by Philip Manning (1992).

One of the key background figures here is Thomas Schelling. Schelling's *The Strategy of Conflict* (1960) was, and in many respects continues to be, an incredibly influential book. The influence, it should be said, was mutual: Goffman's *Strategic Interaction* (1970) was in fact the 'result of a year's sabbatical spent at Harvard with Schelling', while Schelling, in 1986, had said that he 'had been "hooked" on Goffman for thirty years' (Manning 1992, 61). As Manning says, '*The Strategy of Conflict* is a rich source of ideas about the ramifications of game analogies for the interpretation of social behaviour' (Manning 1992, 61). What Manning limits himself to, as we will here, is identifying two features of Schelling's work that were also of central importance for Goffman.

The first feature is that Schelling identified three kinds of games – zero-sum, co-ordination and mixed-motive games – and argued that it was the third that offers the most accurate analogy for social life (though the other two were said to also be useful for certain purposes). Zero-sum games, like chess, are 'governed by "pure conflict"'; co-ordination games, like charades, are games 'where pure conflict is to everyone's disadvantage'; and mixed-motives games are 'where players must reconcile zero-sum ambitions with the possibility of co-operative gains' (Manning 1992, 62). It is instructive that when it came to explaining co-ordination, Schelling made reference to 'traditional sets of rules which people can expect to be conspicuous candidates for general adoption' (Manning 1992, 62). Thus, for example, where it came to explaining 'everyday courtesy', Schelling referred to 'the rules of etiquette and social restraint' – the point, in the context of game theory, being that the 'force' of these rules 'seems to depend on their having become "solutions" to a co-ordination game: everyone expects everyone to expect observance, so that non-observance carries the pain of conspicuousness' (Schelling quoted in Manning 1992, 62). The way that co-ordination is conceived of here (i.e. as a kind of 'tacit social contract' governed by certain 'traditional rules': Manning 1992, 62) is significant, for it shows that despite offering three different kinds of games, Schelling retains the basic picture of an underlying domain or structure composed of rules: either rules that enable pure conflict (as in zero-sum games) or rules that enable co-ordination (as in the rules one expects everyone else to observe in coming to converge on the same solution). This is

important because the mixed-motive game is a mixture of zero-sum and co-ordination elements: again, recall that mixed-motive games are games 'where players must reconcile zero-sum ambitions with the possibility of co-operative gains'. There are unique features, of course, that result from this tension between the two sets of elements, such as a certain amount of uncertainty ('players often find themselves bargaining for resources without knowing the value opposing players attribute to them': Manning 1992, 63), but what is noteworthy for present purposes is that the basic picture of conflict-enabled-by-rules on one side and co-ordination-on-the-basis-of-shared-rules on the other side remains intact.

The second feature of Schelling's work that Manning mentions is his notion of trust. Trust becomes important in mixed-motive games. In zero-sum games it is suicidal, and in co-ordination games it is a prerequisite (Manning 1992, 63), but the key question is how it is imagined in mixed-motive games. The thrust, at least as reported by Manning, of Schelling's view of trust in mixed-motive games was that it represents a commitment to what is likely to be taken for granted: in Schelling's words, 'hardly anything epitomises strategic behaviour in the mixed-motive game as the advantage of being able to adopt a mode of behaviour that the other party will take for granted' (Schelling, quoted in Manning 1992, 64). This, says Manning, is also what Goffman took from Schelling, i.e. for Goffman 'Trust is a necessary condition for routine social life to be possible: without an implicit and unconsidered confidence that buildings are safe, that supermarket food is hygienically packaged, that traffic rules will be obeyed, that pedestrians travel the streets without malign intent, "normal life" would be transformed into unmanageable and shapeless mayhem' (Manning 1992, 64).

The two features identified by Manning in Schelling's work, and the way that Manning articulated the importance of this for Goffman, raises a commonly encountered idea and also problem to be found in sociological extensions of the analogy of games to social life. When we picture persons playing a game, we (as theorists) are confident we can describe them as taking a certain set of fairly well established and stable rules for granted: e.g. when playing chess, a theorist is confident that she can describe the two players as taking for granted the rules for how many pieces there are, how the pieces move, and so on. In other words, the domain (let us stick with chess) is one where it is thought that whatever might arise as a controversy has already been regulated by some rule dictating how that controversy

ought to be resolved, e.g. if one player asserts that the other player made an illegal move, we can look up the rule book to see what counts as making a legal move. When the game analogy is extended, the same stability is carried over to social life – as with the game scenario, so in social life persons are imagined as interacting by expecting each other to observe the same rules (and also mirroring this expectation directed at others by feeling it as an obligation for themselves).[189]

There are many problems with this strategy, and we have already articulated some of them. Let us briefly re-state these problems in this specific context. First, insofar as one might accept this reading of games, it surely seems too strong as a claim for social life in general. This is because social life, unlike games, is not already – and does not come pre-packaged as – one big domain, or a unified set of many domains, in which there are well-established rules for resolving potential disagreements. We cannot foresee what disagreements may arise, which will call for a resolution (and indeed may never be resolved). It is not possible for such potential disagreements to have always and already been prefigured or predicted by any pre-existing rules (e.g. so-called 'implicit rules'). In other words, insofar as one describes persons exclusively as taking for granted certain rules when interacting, one will miss the sense in which participating in social life may involve – and perhaps it may involve it more often than we think – the kind of creativity and negotiation that characterises coming up with resolutions (sometimes in the form of new rules, norms or conventions, and sometimes in other more pragmatic ways, including ways that defer any potential resolution) that do not depend on prior knowledge of pre-existing (e.g. implicitly or tacitly existing) rules. Another way to put this is to say that it is possible but misleading to describe cases in such a way that it looks as if persons who solve certain co-ordination dilemmas (let us adopt this language for a moment) did so, and were always going to do so, on the basis of some pre-existing rule: but what this

---

[189] This two-faced feature of social rules is noted by Goffman. For instance, in *Interaction Ritual* (1967), Goffman says: 'Rules of conduct impinge upon the individual in two general ways: directly, as obligations, establishing how he is morally constrained to conduct himself; indirectly, as expectations, establishing how others are morally bound to act in regard to him. A nurse, for example, has an obligation to follow medical orders in regard to her patients; she has the expectation, on the other hand, that her patients will pliantly co-operate in allowing her to perform these actions upon them' (Goffman 1967, 49). There is some ambiguity in this passage, and some dissonance between the generalisation and the example, for the rules referred to in the example are regulations issued by a medical authority, while the rules in the generalisation are given a moral character. Elsewhere in *Interaction Ritual*, though, Goffman gives examples of what we might more readily think of as social rules, such as rules for how close one stands or sits next to each other (see Goffman 1967, 65-6, including in a mental hospital environment).

misses is that social life is not so closed a domain that all that could ever arise as a dilemma has somehow always and already been prefigured by something that pre-exists it arising (the point is that cases can be described in such a manner that this looks to be so). This is why, for instance, it is misleading to say – as Manning says above – that persons proceed through social life necessarily and exhaustively saddled with expectations, e.g. that supermarket food will be hygienically packaged. This is misleading, for if one describes cases in this way one will miss the fact that it has not always been the case that persons went around expecting this; there arose (or so we can imagine), at some point prior to any such expectation being formed, a problem where some supermarket food led, say, to someone becoming very sick, such that the way the community resolved the dilemma was to establish a rule or norm that the food was to be properly packaged. Putting it as Manning does suggests that persons were always and already equipped with a solution for the dilemma, even before it arose (in other words, the description of the case makes it look like there was always and already a rule waiting to be unearthed that would resolve the dilemma, predicted by the rule, in just that way). What this misses is that persons may have decided the dilemma in a different way (e.g. requiring supermarkets only to sell food that would not be in danger of risk to hygiene). The point is to allow for expectations – understood in the form of the contingently exercised active treating of some behaviour as required – but not to focus exclusively on those expectations, as if persons could be sufficiently describable on the basis of having internalised them and now be following them (by default, unless given reason to do otherwise).

But second, one should not accept a reading of games as domains so closed as the analogy suggests they are, i.e. there are dilemmas that also arise, and that will continue to arise, in games that are also not prefigured by any particular rules, norms or conventions that might, at any one time, be treated as in force and be used to solve dilemmas previously encountered. The problem here is the same: when understood in a certain way (i.e. as constituted by rules), games make it look as if all that could arise as a problem within the game is already solvable by reference to the rules. But consider the following example: at some point in the history of chess, someone complained that the other player's smoking was bothering him; perhaps nothing was done the first time this happened, but over time, these complaints kept on re-occurring; eventually, something was done and a rule was introduced that banned smoking at chess tournaments. There is something to be said for the idea that once

introduced, all chess players that play in chess tournaments share commitment to the expectation that players will observe the rule. But it makes no sense to describe players prior to that rule being introduced as somehow being committed to resolving a potential dispute over smoking in that way (it was not only left unresolved for a while, but it also may have been resolved differently). Further, it would also be inaccurate to describe players in tournaments as somehow being committed to resolving a dispute they knew existed – i.e. they knew that smoking was controversial, but not yet officially banned – by reference to the rule that ended up being introduced; in other words, there would have been a period in the history of chess when this was too uncertain, and where players may have had a variety of views and feelings about smoking in tournament chess. The general point being made here is that it is only when we hold things fixed – as we do when we imagine that persons participate in an activity where all that could arise as a dilemma requiring attention is always and already solvable by some pre-existing rule, norm or convention – that we come to think of trust as something that has as its exclusive objects some rule (or a combination of rules, norms and conventions). This understanding of trust is insufficient, but perhaps even understood without the exclusive focus on taking certain rules for granted, the concept of trust (or taking things for granted) cannot do all the work that it is often put to: even when we are participating in some domain where certain rules are often (though, on our view, contingently) taken as guides and standards for evaluation of conduct (where those rules have been previously introduced as solutions), it is the case that when we participate in that domain we will not (we simply could not) have all the various potential things that could go wrong in our minds (including, or perhaps especially, tacitly). This being impossible is the reason for why it is misleading to describe persons interacting only insofar as they expect each other (and obligate themselves) to observe certain rules, and on that basis, certain specific actions. Persons can treat certain rules (and thus required behaviours) as unchangeable, and can expect persons to comply with those requirements, but this is something built (contingently, and when necessary) on top of an underlying interactive ability, i.e. the interactive ability does not consist in being so guided (by default) by rules that allegedly exhaustively govern some domain of interaction. Trust, one might say, needs to be imagined to be more interactive: not as a commitment to necessarily shared rules, norms or conventions, but as the willingness

and ability to negotiate and cope together with any potential problem that might arise. Having expressed these issues in a general way, let us return to Goffman.

Throughout his corpus, Goffman often referred to social order as composed of rules as analogous to not only the rules of games, but also traffic and grammar rules. Thus, for instance, in *Interaction Ritual* (1967), Goffman says that 'To study face-saving is to study the traffic rules of social interaction; one learns about the code the person adheres to in his movement across the paths and designs of others' (Goffman 1967, 12; Goffman is speaking, in this context, of 'face-work'). Sometimes, the rules are just surface manifestations or instantiations of underlying norms, and it is the norms that are referred to as akin to traffic rules (see Goffman 1963, 8; and Goffman 1971, 26-40; discussed in Manning 1992, 73). This view of the relations between rules and norms is in fact present from very early on: 'Underlying each kind of social order we find a relevant set of social norms… Norms, and the rules in which they are embodied…' (Goffman 1963, 343; quoted in Manning 1992, 73). Statements such as this convinced Manning that 'Goffman characterised social rules as invisible, underlying codes governing our behaviour', where the rules 'simultaneously regulate and constitute the structure of social interaction, usually as background assumptions' (Manning 1992, 73). In *Interaction Ritual*, one finds reference to the same image (orders constituted by rules sometimes described as collected into codes), being said by Goffman to be applicable to a great variety of domains, and not only determining, in each one, what is appropriate and inappropriate, but also in some sense 'guaranteeing' appropriateness: 'In all societies, rules of conduct tend to be organised into codes which guarantee that everyone acts appropriately and receives his [sic] due. In our society, the code which governs substantive rules and substantive expressions comprises our law, morality, and ethics, while the code which governs ceremonial rules and ceremonial expressions is incorporated in what we call etiquette' (Goffman 1967, 55).[190]

---

[190] Goffman is, in this context, speaking of 'ritual order', where he also makes a distinction between 'substantive rules' and 'ceremonial rules'. A substantive rule is one 'which guides conduct in regard to matters felt to have significance in their own right, apart from what the infraction or maintenance of the rule expresses about the selves of the persons involved' (Goffman 1967, 53). 'Do not steal' is an example of a 'substantive rule' (Goffman 1967, 53). A ceremonial rule is one 'which guides conduct in matters felt to have secondary or even no significance in their own right, having their primary importance – officially anyway – as a conventionalised means of communication by which the individual expresses his character or conveys his appreciation of the other participants in the situation' (Goffman 1967, 54), e.g. as when a 'man calmly steps aside to let a strange lady precede him into a lifeboat' (Goffman 1967, 54, fn. 5).

As much as he relied on the analogy with games / traffic rules / grammar rules, Goffman was also careful to distance himself from it somewhat, but it is crucial to see that the way in which he did so was to say that 'even quite formalised codes, such as the one regulating traffic on roads, leaves many matters tacit' (Goffman 1971, 126; quoted in Manning 1992, 74)[191] – the idea, then, nevertheless being that where the explicit rules run out there are implicit rules (or implicit norms available to be articulated into rules). This appeal to implicit rules has two effects: first, it turns games / driving / speaking a language into domains governed by – as Manning puts it – 'rules [that] are fixed and determinate', and second, it pictures social life as nevertheless always and necessarily governed by rules, even if more implicit and more indeterminate ones (see Manning 1992, 77). The general effect of this is to extend, rather than let us see the limits of, the governance view.

There is quite some ambiguity and confusion in how Manning presents Goffman's view of rules. For a start, there are many different terms that Manning refers to as having been used by Goffman to make similar points. Thus, for instance, consider the following statement (which refers not only to different kinds of rules, but also to 'practical knowledge' and 'principles'): 'Rules are neither mere regularities nor are they laws. They are the practical knowledge of how to carry on; to follow a rule means to apply principles to circumstances in an indeterminate manner' (Manning 1992, 77). On this way of putting things, Goffman did not identify the rules – perhaps he presupposed them – instead, what he identified were the 'general principles to which we appeal in following everyday rules' (Manning 1992, 77). Adding to the confusion is that these 'general principles' are also at times referred to by Manning as 'constitutive rules': e.g. 'At a great level of abstraction [Goffman] claimed to have identified the constitutive rules which pervade every face-to-face interaction', calling them also 'general guidelines about the kinds of things that happen during face-to-face interaction', as well as 'the "syntax" of everyday behaviour, without which the "language" of behaviour is incomprehensible"' (Manning 1992, 77-78). There is a mixture of analogies and terms here, which makes

---

[191] Although, in *Interactional Ritual* (1967), Goffman focuses on 'ceremonial rules' (as distinct from 'substantive rules'), it is clear from his analysis and description of cases that these are implicit or tacit rules, revealed in persons' gestures and facial expressions.

it difficult to reconstruct Goffman's view. Yet more terms are introduced in the following passage:

> Goffman thought that social rules are largely constitutive of what we do when we carry on in a way of life: they form a deep structural code which generates an infinite variety of surface manifestations… We use them to make sense of the stream of face-to-face encounters. To understand a community, a lifeworld, a form of life, is in part to understand how perceived events correspond to this underlying code… Goffman formulated a small number of interactional principles that loosely regulate interaction; they are schemata of interpretation. One of his most persistent beliefs is that for behaviour to be construed as understandable – indeed, as sane – it must be generally acknowledged to have been derived from a set of interactional principles. (Manning 1992, 78)

Perhaps the best we can say on the basis of Manning's reconstruction, is that Goffman positions persons as acting and interacting in contexts where certain rules for interaction are assumed to apply (to be relevant), where some of those rules may be explicit and some may be implicit, and where both the explicit and implicit ones may be understood as surface manifestations of more general norms, such that what Goffman did in analysing persons positioned in this way was to identify general principles for acting and interacting in contexts constituted and regulated by such rules (as instantiations of more general norms) – but all this with the proviso that we retain a very broad understanding of what a 'rule' might be, including appropriate body postures, spatial arrangements, conversational courtesies, etiquette and so on (Manning 1992, 79).

On this reconstruction, then, the general idea is that persons are pictured as acting and interacting in quite controlled environments, where what is appropriate and what is not is defined (indeed, appropriateness and inappropriateness being so defined is also what constitutes, brings into existence, the very environments in which persons are described as interacting), and where persons expect each other or assume each other to be committed to those ways (in which what is appropriate and what is not is defined), such that if persons display either unwillingness or inability to conform to those rules, then they may be shunned by other members of the group or – in environments such as asylums – they may be evaluated more formally and possibly

punished by an authoritative, policing group.[192] In reply to this we should say that we can agree that there can be environments (so-called 'heavily regulated environments') where persons actively treat certain rules and their requirements as unchangeable, but then it is vital that when we study these environments we do not present persons as but always recognising and re-enacting them, but where we see such active treating as contingent attitudes built occasionally on top of interactive abilities. That there is a need for such an underlying interactive ability becomes clearer when one identifies gaps between any such rules and how the policing group exercises its evaluation – e.g. 'turning a blind eye' – and the little 'resistances' persons introduce and experience, even or perhaps especially in such environments.[193]

It may very well be that Goffman can be re-read from a second-person perspective, where the focus becomes not on using the many gestures and expressions he identified in, for instance, face-to-face work, as revealing of an underlying order (and / or the result of a projected / expected order), but rather as part of the balance and imbalance of precisely imperfect (always somewhat rough, always somewhat askew) interactions between persons. *Interactional Ritual* (1967) is, after all, full of fascinating descriptions of interpersonal dynamics (e.g. subtle modifications in intonation), but these are always subjected by Goffman to a characterisation of interaction according to rules (e.g. the exact kind of intonation is responsive to rule-defined deference).[194] The other difficulty is that Goffman's social agents are often depicted as very strategic, calculating in advance, sometimes explicitly and sometimes tacitly, what they ought to do in order to appear to others in certain

---

[192] In *Interaction Ritual*, Goffman says: 'Much psychotic behaviour is, in the first instance, a failure to abide by rules established for the conduct of face-to-face interaction – rules established, that is, or at least enforced, by some evaluating, judging or policing group' (Goffman 1967, 141; see Manning 1992, 79; see also Goffman 1961).

[193] Indeed, it could be that at least some forms of psychotic behaviour are precisely the inability to show and experience resistance and independence, and thus abiding too much by rules (though just not the ones that the policing group requires). Michel de Certeau's work (1984) is famous for its appeal to everyday resistance. It needs to be added, however, that one has to be careful about how one conceives of 'resistances', e.g. if persons are described as but resisting some allegedly always and already established order of rules, then one precisely presupposes the governance view (the point is that to get rid of the governance view one has to avoid the idea that persons are either always and necessarily acting in accordance with or violating conventions, norms and / or rules).

[194] For example, in *Interaction Ritual*, Goffman describes rules of advice given to 'masters' on what tone of voice (including when to intonate) to use when addressing 'servants' (see Goffman 1967, 62). In many of these examples, Goffman relies on etiquette books, which is somewhat problematic for he then assumes certain rules (as described in those books) are relevant, and describes the gestures and expressions of persons as if they were committed to them.

ways.[195] The problem with this is that it encourages one to think of the body as passively controlled by a mind, the activity of which is limited to orchestrating events (including their own gestures and expressions) in accordance with already mastered rules. This not only appears to be committed to a kind of body-mind dualism – where the body is perpetually passive and doomed to being directed by a mind – but it also greatly narrows how we can describe the mind-body learning and working (which, as a result of presupposed rules, is always described as learning those rules and re-enacting them). On the way we are trying to understand the normative dimension of social life, persons can perform actions when relating with others (that others may experience as appropriate) without having previously committed / obligated themselves to doing those specific actions, and this is 1) partly because the abilities of persons to relate with each other more or less successfully are not describable as rule-structured (we learn and exercise our abilities in more flexible ways than that), and 2) partly because what may be appropriate to do in the course of the relation is simply not exhaustively predictable by some previously adopted rules. There is, on the view argued for here, a basic ability to interact with others – to attend to others, be surprised by them, and be responsive to their actions as unfamiliar invitations and opportunities to relate. The point being made here is that, keeping in mind the above-mentioned difficulties, Goffman's work may be re-readable with a view to seeing that ability, and not some underlying order of rules, as basic.[196]

We have relied, in the above, to a large extent on Manning's reconstruction of Goffman's view, based on a reconstruction of his entire oeuvre, but it will be important not to take Manning's word for granted, and ourselves examine Goffman's work in more detail and on its own terms. We will do so by reference to perhaps the best, but also most complex, summary that Goffman ever gave of his work: namely, his 1982 Presidential Address to the American Sociological Association entitled 'The Interactive Order' (1983).

---

[195] Though it is also by no means the case that social agents are depicted by Goffman as only concerned with how their own face will appear to others, but also as concerned with 'saving' the face of others (see chapter 1 of *Interaction Ritual*, Goffman 1967). Goffman's descriptions of this concern with the face of the other may offer valuable resources for a second-person perspective on the normative dimension of social life.

[196] More of this possibility of re-reading Goffman is said in the Section below (18.B.).

B. The Interactive Order

The general definition of interactive order is of a domain of face-to-face interaction, or body-to-body relation (though including interactions without physical presence, such as those via telephone and mail; Goffman 1983, 2). It is important to see that advocating for this sense of interactive order – and its significance – was part of a general push by Goffman against imposing certain macro structures on the analysis of interacting, as would be the case if one began with 'the standard contrast between village life and city life, between domestic settings and public ones, between intimate, long-standing relations and fleeting impersonal ones' (Goffman 1983, 2). This reference to the importance of studying the micro – of prioritising the micro in analysing social life – is important for this thesis, but the key question is just how one understands interaction and whether one pictures it as occurring in the context of a pre-configured, laid out, already organised order. Goffman's view has elements of both: he clearly wants to take interactive dynamics seriously, but he also wants to position the exercise of interactive abilities within an order imagined to be constituted and regulated by rules (and, at an even more general, underlying, level, norms).

Goffman begins by outlining some of the specific 'processes and structures' of the interaction order (Goffman 1983, 3). One of these is that the interaction in question has to happen within a relatively circumscribed space and time. This may be the result, for instance, of having to use the same 'fixed equipment' (Goffman 1983, 3). The general point is that there may be 'all kinds of unsentimental and uninherited reasons why individuals everywhere – strangers or intimates – find it expedient to spend time in one another's immediate presence' (Goffman 1983, 3).

The next step is a crucial one: once in each other's physical presence, Goffman argues that persons both present themselves and can view others as presenting themselves in terms of standardised bodily and vocal behaviour. This standardisation (which Goffman also refers to as 'social ritualisations') enables 'the sustained, intimate co-ordination of action, whether in support of closely collaborative tasks or as a means of accommodating closely adjacent ones' (Goffman 1983, 3). Goffman presents another person's 'intent and purpose' as being directly available to observation, and this he attributes to two different origins: 'the categoric kind involving placing the other in one or more social categories, and the individual kind, whereby the subject under observation is locked to [sic] a uniquely distinguishing identity through appearance, tone of voice, mention of name or other person-

differentiating device' (Goffman 1983, 3). Persons who face each other are also vulnerable in all kinds of ways – they are vulnerable, for instance, to 'physical assault, sexual molestation, kidnapping, robbery and obstruction of movement' – and they may also bond together (rather than hurt each other; Goffman 1983, 4).

When we confront another person, the ways in which make them intelligible to ourselves (via the two modes of observation noted above), detecting either or both 'enablements and risks' will depend on what kind of 'situation' we find ourselves in (Goffman 1983, 4). This 'situationism' (as Goffman himself calls it: 1983, 4) is another crucial element in the view: 'it is in social situations that these enablements and risks are faced and will have their initial effect. And it is social situations that provide the natural theatre in which all bodily displays are enacted and in which all bodily displays are read' (Goffman 1983, 4).

Let us keep note of the language here: there are 'situations' that persons 'face' (Goffman refers to this as 'confrontational', saying 'our experience of the world has a confrontational character': 1983, 4), and in those situations persons 'enact' certain 'displays' which are 'read' (or also 'observed') by others, thereby making each other intelligible to themselves (such as to detect whether there is a risk to be defended against or a bond to be built). Explanatory pressure here comes from two different directions: first, from the direction of situations 'that sustain an interaction order' (by 'situations' Goffman appears to include: the local bar, a small shop floor, a domestic kitchen: see 1983, 4); and second, from 'the established biography of prior dealings with the other participants' brought to the situation by an individual (Goffman 1983, 4). This tension is never quite resolved in Goffman – indeed, the way he treats it suggests he would welcome that tension – but what is common to both is that something must be (by each of the participants) assumed to be shared, whether it be so assumed because of the 'situation' (and, one presumes, on this basis the ascription of a 'category' to the persons – e.g. a waiter in café) or because of some biographical history unique to that person.

Goffman's next move is to further elaborate on what he means by the 'order' aspect of 'interaction order'. We must proceed carefully here, for Goffman is himself careful to protect his view against what he calls certain 'dogmas' about order. His general view is that 'as an order of activity, the interaction one, more than any other perhaps, is in fact orderly, and…this orderliness is predicated on a large base of shared cognitive presuppositions, if not normative ones, and self-sustained restraints'

(Goffman 1983, 5). He further elaborates on this by saying that 'the workings of the interaction order can easily be viewed as the consequences of systems of enabling conventions, in the sense of ground rules for a game, the provisions of a tax code, or the rules of syntax of a language' (Goffman 1983, 5), but he immediately distances this from two ways of reading it that would be too dogmatic. The first of these is a 'social contract' view, according to which 'any convention that facilitates coordination would do – [where] the several conventions [alternatives to each other] in themselves have no intrinsic value' (Goffman 1983, 5). The second is the 'social consensus' view, according to which 'orderly interaction is seen as a product of normative consensus, the traditional sociological view that individuals unthinkingly take for granted rules they nonetheless feel are intrinsically just' (Goffman 1983, 5). Goffman criticises these two dogmas, but only in the limited sense that they over-simplify the motivations that persons may have for following the rules: 'Individuals', he says, 'go along with current interaction arrangements for a wide variety of reasons, and one cannot read from their apparent tacit support of an arrangement that they would, for example, resent or resist its change' (Goffman 1983, 5). This is an important qualification, but it nevertheless keeps what is common to both, namely that social order consists of the acceptance of 'conventions and norms as given', and that it this social order (a kind of syntax, as we saw above) that 'allow[s] a great diversity of projects and intents to be realised through unthinking recourse to procedural forms' (Goffman 1983, 6).[197]

There is some confusion here when Goffman says that these 'ground rules inform the interaction order and allow for a traffic of use' (Goffman 1983, 6), because for Goffman it is not so much a question of 'informing' the interaction order as constituting it (it is order insofar as there are conventions and norms that are taken for granted). Either way, the basic picture is of persons interacting within such an order, which, to make it compatible with what was said above, manifests itself in distinct

---

[197] Nothing in this thesis is an argument against the ability of rules to enable creativity; e.g. it is acknowledged that the use of constraints for this purpose is very well-known in the history of literature (indeed, the French movement *Ouvroir de littérature potentielle* (*Oulipo*), literally, 'the workshop of potential literature', was based on it; for a discussion, see Andrews 2003). The point is not to confine, for instance the activity of literary writing, to a matter of acting in accordance or resisting such constraints (as much as reference to such constraints can help us understand some aspects of certain works). We do not have to be following or violating some pre-existing rules (or any other allegedly implicitly or tacitly existing constraints) to be creative. Of course, this is not a thesis about 'creativity'; where it appeals to creativity is in contrast to the passivity involved in depicting persons as but re-enacting or violating pre-existing rules.

situations that persons who interact confront, such that persons detect the kind of situation they are faced with, categorise others in certain ways they have learnt to categorise them, and then, taking the conventions and norms of the order for granted, enact and display (and read the other as displaying) the appropriate forms of action and interaction that conventions and norms establish.

Aside from the above general account of the 'processes and structures' of Goffman's notion of interaction order, Goffman also provides examples of 'the basic substantive units, the recurrent structures and their attendant processes' belonging to interaction order (Goffman 1983, 6). Here, Goffman speaks of persons as 'ambulatory units', either 'singles' or 'withs' (i.e. more than one), such as 'files and processions' and queues – the general idea being that these can be 'treated as self-contained units for the purposes of participation in the flow of pedestrian social life' (Goffman 1983, 6).[198] There are variety of ways in which such units can interact, including in momentary contacts (such as passing glances), or in conversational encounters that may include formal meetings (as in the case of legal proceedings), or in less conversational encounters as in 'card games, service transactions, bouts of love making and commensalism' (Goffman 1983, 7). Any such encounters between units, Goffman continues, may happen on a 'platform' where an 'activity is set before an audience', as may be the case for instance with 'a talk, a contest, a formal meeting, a play, a movie', etc (Goffman 1983, 7). These encounters between units, which may occur on a platform (watched by others), may all form part of a 'social occasion', by which Goffman means 'the foregathering of individuals admitted on a controlled basis, the whole occurring under the auspices of, and in honour of, some jointly appreciated circumstances' (Goffman 1983, 7). In other words, such an occasion provides 'a setting for many different small focused undertakings, conversational and otherwise, and very often will highlight (and embed) a platform activity' (Goffman 1983, 7). The overall view, then, is of the elements of interaction order consisting of units that encounter each other in all kinds of settings (some conversational, some formal etc) that may occur on a platform and that may be part of an occasion.

Having given this account of interaction orders in general (including basic elements of them that can be used in analysing instances of interaction), Goffman goes on to consider the relationship between interaction order and macrosociological

---

[198] The reference to units and flows is designed to echo the previous references to 'traffic rules' that enable these units to participate in these flows.

explanatory devices, such as social structures, relationships and statuses. We cannot examine this in detail here, but it will be useful to consider some aspects of Goffman's discussion.

At each turn in the discussion, Goffman pulls apart encounters that occur in situations with more 'traditionally considered elements of social organisation' (Goffman 1983, 8). He speaks, for instance, of 'institutionalised settings' including everything from 'states to households' as well as of 'small structures' such as 'corner businesses, families, relationships' (Goffman 1983, 8). Goffman is careful about over-simplifying the relations between these situationally embedded encounters and structures – e.g. he finds 'uncongenial' claims that 'all macrosociological features of society, along with society itself, are an intermittently existing composite of what can be traced back to the reality of encounters' (Goffman 1983, 8), and he asserts that he does not 'believe that one can learn about the shape of the commodities market…or the structure of kinship systems…from particular social encounters among the persons involved in any one of these patterns' (Goffman 1983, 9).

What makes this difficult to assess is how exactly Goffman would articulate the difference between situations (or what he also sometimes refers to as 'settings') and structures. Thus, for example, Goffman says that there is a '"loose coupling"…between interactional practices and social structures, a collapsing of strata and structures into broader categories, the categories themselves not corresponding one-to-one to anything in the structural world, a gearing as it were of various structures into interactional cogs' (Goffman 1983, 11). He also adds that there is no 'set of transformation rules, or a membrane selecting how various externally relevant social distinctions will be managed within the interaction' (Goffman 1983, 11). These are promising statements, but the problem is that in both these cases Goffman has given us no independent way of understanding what he means by 'interactional practices', for he has described these as enabled by ground rules and categories – i.e. because he described persons as interacting only in the sense in which they share certain background assumptions (the categories, as well as the norms and conventions) and enact the various displays and readings these enable, it makes it difficult to see how 'interactional practices' (or 'encounters' in situations / settings / occasions) is at all different from various ways of articulating structures (whether these be specific organisations in which certain roles are played out, or certain cultures in which certain statuses are assigned, etc). In the abstract, then, it seems very

reasonable to assert that there are no 'transformation rules' or some absolute 'membrane' that dictates what is relevant when, but in the absence of a description of instances of interaction as ones where relevance remains indeterminate and prospective (i.e. not determined in advance), it is difficult to see how although he may not explicitly posit such transformation rules, Goffman nevertheless presumes or presupposes them.

In case this is unclear, let us try to articulate it in another way. If we think of interactions as occurring in situations in which persons display themselves, and are read by others, in accordance with certain categories as these are attached (in some way) to the ground rules (i.e. the norms and conventions) that are inherent in the situations (for they enable the interaction in the first place), where is there room for a distinction between 'situations' and 'structures'? Perhaps what one could say is that Goffman is resisting the pull of certain traditional ways of conceiving of structures, and that is a fair response, but on a general level, is he introducing anything different? Is he not replacing very large and general structures with simply more small-scale and minor structures, which he calls situations (or settings / occasions, etc)? What is common to these terms – i.e. to structures, settings, situations, and occasions[199] – is that they are at once ontological and epistemological, i.e. they are at once the elements that make up social knowledge as well as the social contexts in which persons are described as interacting. There seems to be a perfect fit between, on the one hand, the situations in which persons encounter one another, and the resources that they need to interact with one another. Persons interact in structures, settings, situations, occasions or cultures, and what enables them to interact with one another is that they share the background assumptions that constitute these very structures, settings, situations, occasions or cultures. It is in these contexts that persons display themselves and read each other, and it is only in those contexts that they could do so. The ground rules – the norms and conventions – enable the interaction, and it is knowledge of those ground rules that persons carry around with them and enact (instantiate) when they do interact with one another.

What makes one think there is a difference here between situations and structures is that Goffman refers to structures as general explanatory schemes, and

---

[199] Goffman also refers at one point to 'cultures' which, he says, 'have a vast lore of fact and fantasy regarding embodied indicators of status and character, thus appearing to render persons readable' (Goffman 1983, 8).

what he (very reasonably) points out in his paper is how one ought not to assume that those explanatory schemes will always be relevant. Consider the following examples:

> The precedence one gives to one's immediate boss one gives to his or her immediate boss too, and so on to the head of the organisation; for precedence is an interactional resources that speaks to ordinal ranking, not to the distance between the rungs. (Goffman 1983, 12)

> Thus, in recent times blacks and women have concertedly breached segregated public places, in many cases with lasting consequence for access arrangements, but, all in all, without much change in the place of blacks and women in the social structure. (Goffman 1983, 12)

What these examples indicate is that the way Goffman uses the concept of 'structure' is as a specific kind of explanatory scheme, e.g. a scheme of organisational hierarchy, or a scheme based on the categories of 'blacks' and 'women', and what he wants to say, by reference to the notion of a 'situation', is that when one describes interaction in a situation one ought not to assume that the explanatory scheme itself will tell you (the theorist describing the scene) what is going on, for participants may be using other 'interactional resources' or doing things that 'breach' the actions associated with the categories of the scheme. This is a valuable observation, for it points to the limits of any explanatory scheme (i.e. the limits of traditionally relied on 'structures' to account for social life), but the point being made here is that the same applies to situations (and all the various categories, norms and conventions that these are said to be composed of, and which are said to enable interaction).

Goffman rightly points out the explanatory limits of structures, but he seems to make the same mistake he accuses them of making, i.e. to neglect to see the explanatory limits of that which he uses to describe interaction, i.e. situations conceived of as containing ground rules connected to shared cognitive presuppositions that enable interaction. But just as those who work with, say, an organisational structural explanatory scheme would miss the sense in which someone can give the same amount of precedence to persons many rungs apart on the organisational ladder, so someone working with an explanatory scheme based on to whom one gives precedence (i.e. someone who describes persons as sharing the cognitive presupposition to give precedence to one's superior) may miss the sense in which someone may be subtly undermining or ridiculing the boss or the boss's boss, and so on. The point is a simple one: Goffman's 'situations' (or, put differently, his

traffic rules, or his syntax) have no more special metaphysical status than the various structures used traditionally to explain social life (i.e. to establish the sense in which it is ordered and then picture persons as socialised into that order such that they re-enact it on single occasions). Of course, Goffman is right that those traditional structures may be wildly outdated – or if not wildly outdated, that they will, say, miss out on passing 'fads and fashions' (see Goffman 1983, 12) – but none of this means that his own way of articulating the situations in which persons act does not have limitations of its own. Goffman offers us explanatory schemes that we find resonate with us more than certain traditional schemes do (though the above reference to 'blacks' for instance now reads awkwardly – to say the least – to our ears), but this does not make his reference to situations (hallways, shop room floors, domestic kitchens, etc) any less of an explanatory scheme. The common point is that all these explanatory schemes put persons in contexts where those contexts are assumed to hold still what is relevant, such that participation in those contexts becomes possible only by internalising what the contexts require to be taken into account (our point being that relevance cannot be held still, or simply deemed to be pre-determined, like this).

You might ask: but whether we speak of situations or structures, what else do we have? How else would you explain social order? And here we come to one of the key points: why presuppose that the explanatory target is social order, i.e. the repetition of certain kinds of patterns? What happens when we presuppose this is that we first identify the patterns as instantiations of certain rules, and then we picture persons as learning those rules, such that we then describe everything that they do as but instantiations – enactments – of those rules (i.e. we describe persons' behaviour as patterned precisely in the way that our rules predicted it). Then, when we come across behaviours that we find it difficult to see as instantiations of rules, we do not change the scheme but rather say that the scheme includes inherently (implicitly, tacitly) some rule that does make sense of that behaviour after all. The result is a kind of self-fulfilling prophecy in which the description offered always and unerringly gives evidence of the sufficiency of the explanatory scheme. The explanation works by postulating that all that could go wrong has been pre-figured or predicted by the rules (for even if one comes across something that is not so easily explicable, one posits the existence of some implicit rule that prefigures or predicts it after all), but what the explanation fails to see is that there are many matters, which we do not and cannot have knowledge of in advance, that could go wrong if only because it is us who need

to (in the sense of are able to, and not must there and then) decide whether to see something as just an instantiation of a rule or as something else – as something potentially requiring a new rule (or perhaps something left unresolved).

It might help to put this in terms of an experience one can have: thus, one can have the experience of expecting someone to do such-and-such a thing, such that when they do what one does not expect them to do, one evaluates them negatively (as breaching that expectation), but one can also have many other experiences with others where one does not assume that they will do what one expects to happen. It is also possible that when faced with something one did not expect someone to do one reacts not by reinforcing one's expectation and questioning the behaviour of the other, but by questioning one's expectation, e.g. by asking oneself: 'perhaps what I thought was relevant / important is not (at least not here, not now, not with this person)?' The point here is that which way one goes is not pre-determined, but that no room is left open for indeterminacy when one's explanatory target is social order – for if one presupposes a social order, all that is done is either in accordance (often assumed to be positive) or in breach (often assumed to be negative) with the explanatory structure that established / defines that order (and where every act but reveals some allegedly underlying enabling rule / pattern). Social 'order' is too rigid, too demanding an explanatory object; in presupposing repetition of certain kinds of action, and making the re-enactment of those same actions a condition of possibility for interaction, it closes off – and does not allow us to see – the un-patterned (but not chaotic) ways in which we relate with each other.[200]

C. Embedded Encounters: Echoes of Goffman

Goffman's work has been highly influential in shaping the way contemporary sociologists analyse face-to-face encounters. A relatively recent example – and one that fuses together many of the strands of this approach – is Jonathon H. Turner's *Face to Face: Toward a Sociological Theory of Interpersonal Behaviour* (2002). Turner draws heavily on Goffman, arguing that his book builds on the way in which Goffman suggested encounters were embedded in 'social occasions' (Turner 2002, 32). Turner's interpretation of what Goffman meant by 'social occasions' is that they are dictated by meso- and macrostructures (Turner 2002, 32; see also 21-27), which

---

[200] Again, one has to avoid the dichotomy here, so even saying 'un-patterned' is already too strong: the point it is a mix of falling in and out of balance – more on this in Part V.

further supports the reading of Goffman's 'The Interactive Order' (1983) given above. In other words, on Turner's reconstruction, Goffman's social occasions are crossroads for potentially many overlapping structures, but the effect is still the same: persons are still positioned in contexts where certain structural elements are thought to be present and determinative of how the encounter unfolds.

Turner argues that all encounters are embedded in meso- and macrostructures, which include what he calls corporate units (e.g. multinational corporations, governmental bureaucracies, but also various internal partitions of these and other kinds of collective organisations that have divisions of labour) and categoric units (e.g. gender, age, ethnicity), the point being that, according to Turner, 'what transpires in an encounter is very much constrained by the nature of corporate units' (Turner 2002, 35) as well as categoric units (e.g. 'an encounter among members of the same category (men, for example) will have very different dynamics than one where heterogeneity exists (men, women)': Turner 2002, 37).

One can see how things can be taken too far here: it is one thing to assert that various ways of classifying social reality (e.g. into organisations, and internal divisions, or into classes such as gender, age etc) may be actively treated by someone as relevant to some encounter, but quite another thing to say that such-and-such a person belongs to such-and-such a category and thus will behave in such-and-such a way (the first leaves room for contingency and the limits of explanatory schemes, while the second is deterministic precisely as a result of beginning with and prioritising – indeed, keeping fixed – the alleged buildings blocks of social reality).

Turner, it should be added does pay attention to what he calls 'microlevel reality', but it is significant that these too read like paths set down for interacting with others. Turner speaks, for instance, of types of emotions required to be displayed to others, patterns associated with roles and statuses, and 'symbolic properties' in which he includes scripts. For example, he says: 'All interactions are directed by cultural scripts imposed by the meso- and macrostructures in which an encounter is embedded… When individuals interact, they use cultural scripts and status positions of meso- and macrostructures to signal and interpret each other's gestures' (Turner 2002, 40). Turner's general explanatory strategy, then, is to first outline what the elements are of meso- and macrostructures, then to show how these are enacted through microlevel processes, and then to depict encounters as subject to such

processes in situations that are the meso- and macrostructural contexts in which encounters are embedded.

The point we have been making in the above discussion of Goffman, but also in earlier Chapters of this Part, is that we risk over-generalisation of the contingent usefulness of certain explanatory schemes (just as we do with normative resources) when we transform them into an underlying default order; doing so results in a kind of social determinacy and a narrowing of how minds-in-interaction learn and work (in other words, it makes minds the mere carriers of elements identified by the theorist as comprising social reality and then assumed to always be relevant any encounter).

D. Breaching Experiments

What has been argued in this Chapter applies also to the claims sometimes made on behalf of the results of so-called 'breaching experiments'. In one example of these experiments, Stanley Milgram asked a class of his students to go up to someone on the subway in New York and 'simply ask for his seat' (Milgram 1977, 4). Out of a large class of graduate students, one student, Ira Goodman, volunteered. Goodman asked fourteen people in the subway to relinquish their seat for him – providing no excuse or justification – half of which, to his surprise, got up (he had in fact been asked by Milgram to ask twenty, but reported as being unable to ask more than the fourteen he did ask). Milgram himself tried the task and reported finding it extremely difficult:

> Frankly, despite Goodman's initial experience, I assumed it would be easy. I approached a seated passenger and was about to utter the magical phrase. But the words seemed lodged in my trachea and would simply not emerge. I stood there frozen, then retreated, the mission unfulfilled. My student observer urged me to try again, but I was overwhelmed by paralyzing inhibition. I argued to myself: 'What kind of craven coward are you? You told your class to do it. How can you go back to them without carrying out your own assignment?' Finally, after several unsuccessful tries, I went up to a passenger and choked out the request, 'Excuse me sir, may I have your seat?' A moment of stark anomic panic overcame me. But the man got right up and gave me the seat. A second blow was yet to come. Taking the man's seat, I was overwhelmed by the need to behave in a way that would justify my request. My head sank between my knees, and I could feel my face blanching. I was not role-playing. I actually felt as if I were going to perish. Then the third discovery: as soon as I got off the train, at the next station, of the tension disappeared. (Milgram 1977, 5)

Similar experiments were also performed by Harold Garfinkel. Here is an example Garfinkel reports from student accounts:

> The subject was telling the experimenter, a member of the subject's car pool, about having had a flat tire while going to work the previous day.
> S (Subject): I had a flat tire.
> E (Experimenter): What do you mean, you had a flat tire?
> She appeared momentarily stunned. Then she answered in a hostile way: What do you mean, 'What do you mean?' A flat tire is a flat tire. That is what I meant. Nothing special. What a crazy question.
> (Garfinkel 1967, 42)

Milgram and Garfinkel argued that these experiments reveal 'the enormous inhibitory anxiety that ordinarily prevents us from breaching social norms' (Milgram 1977, 4) or 'the stable social structures of everyday activities' (Garfinkel 1967, 37), but this is not the only way to read these experiments. Certainly, if one sets the experiment up in such a way that one thinks one is targeting a certain norm, then one will explain the feelings subjects might experience in conducting the experiment as anxieties to do with breaching those norms (we have here a perfect example of a self-fulfilling prophecy, where the descriptions are said to be evidence of the pre-supposed explanatory scheme). But why assume that the feelings of anxiety in the subway case or in the case of the flat tire are evidence of pre-existing social norms or stable social structures? Could the feelings – for instance in the subway case – be explained not on the basis that one was breaching some kind of implicit norm (not to ask someone to stand up) but rather, for example, on the basis that one was engaging in very self-consciously artificial encounters with others? In other words, in these experiments, the subjects who performed the experiments were using those others for certain ends where the interaction itself was just a means to that end. More strongly, you might say that the results of the subway experiment – half of the people Goodman asked stood up, and when Milgram tried the experiment the person who he asked also stood up – show that there is no such norm, i.e. that if you ask people to stand up they may or may not (perhaps more likely they will) stand up (perhaps to avoid any kind of conflict or confrontation – the point being that such avoidance is not necessarily explicable on the basis of some underlying norm).

Notice, for instance, that, as reported, the experiments did not consider the experience of those who were asked to stand up: one wonders, then, whether those

persons, who were not primed for a self-conscious exercise of using the other for the purpose of a sociological experiment, suffered the same corresponding feelings (e.g. indignation at having been asked) as the subjects performing the experiment (instead, those persons may have experienced / expressed concern for the person asking; but even if they did experience / express indignation, there is an open question as to whether the feeling would need to be described on the basis of some judgement as to a norm having been broken).

Sometimes, these experiments are very provocative (in a way, the flat tire example is like this – deliberately provoking a response from the other), but the subway case is a good example of how there may not be any stable structure underlying interaction for what persons do (whether they stand up, how they react) is not pre-determined (persons cannot, then, be assumed to share commitment to the same allegedly underlying norm). In this respect, the flat tire example is a strange one: is it being suggested by Garfinkel that that is how we can expect everyone to react, i.e. first by being stunned and then shouting in hostile fashion?

It is important to see how ambitious are the claims that social interaction occurs only insofar as, for it is enabled by, shared commitment to certain social norms, or shared cognitive presuppositions whose content is some set of background assumptions about how others ought to act. According to this line of explanation, anything that we could and do experience in social life is somehow already built into us such that we will respond in the way that stock of expectations allegedly disposes us to respond. But is it plausible to think that the way we respond is only ever evidence of the way we were always and already disposed to respond (where the content of those dispositions is said to be some rule, norm or convention that governs the situation in which we are allegedly interacting)? To think this is to think that what is relevant – what could matter to us, what could be important to us, the way we might care about something – is always and already laid out for us in 'the situation' in which we interact. And to think that is to fail to see how not only do many of us care about different things, but also how we can come to care about things we previously had not cared about, including creating new objects of concern that may have never occurred to us (as a group, say) collectively before: consider, for instance, how we began (not all of us, and not to the same degree, but certainly more than before) to care for the environment (and thus in a way also for future generations). Picturing persons as interacting only insofar as they share commitments to certain rules presupposes that

they care, and will always care, for the same things, but this neglects to see that they may care for different things (e.g. the person in the organisation might not display, or be motivated to display, any kind of precedence for his boss, and nor might the boss expect to be shown that precedence, for the employee and the boss might both not care about hierarchical statuses) and that they may come to care for things no-one else cares for or has cared for (imagine we start caring for the effects of our behaviour on earth on the moon – we cannot at this point imagine why we might do so, but one can imagine the possibility). The point is that theories of social order – whether they appeal to practices or games – tend to set things up in such a way that it becomes difficult to see how any of us could react differently, could care differently, than the rules, norms or conventions that allegedly constitute and regulate the social order predict we will.

Our conclusion should not be that we live in a void where we never expect anything (though this does not mean our expectations must always be shared), or where we never have recourse to normative resources to guide ourselves and to evaluate ourselves and others; nor ought we to discount the force of pressures to converge (we see this force also in the explanations we get used to drawing on). Instead, our conclusion should be that any explanatory scheme that prioritises such resources and that describes our expectations by reference to them (as if we internalised them) has serious limitations. Acknowledging these limitations also means we should avoid, or at least be very careful about, talk of implicit / hidden / invisible rules, norms or conventions – the point being that our postulation of such implicit etc normative phenomena points more to how we ourselves would react and expect others to react, or what we care about, and this is not necessarily generalisable (and thus also not necessarily shared). In short: there is no argument here for the necessity of sharing determinants, just the contingency that we may on any particular occasion actively treat the same normative resources (assuming we also understand them to require the same things) as relevant and applicable.

Chapter 19. Socially Appropriate, Yet Unreflective

We mentioned in an earlier Chapter (13) when examining the recent work of Rietveld, that Rietveld has collaborated with a number of colleagues (Klaassen et al 2010) to extend his thesis of how we can act adequately, yet unreflectively, to the social domain. Examining this account will offer a neat introduction to the material covered in this Chapter. Having examined that account we will then turn to other versions of theories where background abilities, emotions, and dispositions (amongst others) have been offered as carriers (for want of a better term) of our internalisation and sensitivity to necessarily shared determinants of what is socially appropriate and inappropriate. The four versions examined here are: first, John Searle's appeal to the background; second, Pierre Bourdieu's notion of the habitus; third, Hans Joas's articulation of the 'creativity' of social action; and fourth, a recent account of enactive social cognition that does not (as might be hoped from the enactive strategy) move us any further away from the governance view.

A. The Case of Mary and Andra

Recall that for Rietveld the phenomenon he wants to explain is how we can act adequately, yet unreflectively or without deliberation. In his collaboration with Pim Klaassen and Julien Topal (2010), Rietveld extends this object of explanation to everyday life.

Klaassen et al do not define what they mean by 'everyday life', but it seems that the way they use the concept it is meant to extend out from beyond examples of expertise, and thus point to ordinary interactions between persons and their environments, as well as ordinary interactions between persons. Helpfully, they offer an example of such unreflective, yet adequate social action, and it will be profitable to consider their description and explanation. Here is the example:

> Take the situation that Mary is having an intimate yet cheerful conversation with her friend Lucas, and that they are expecting to meet a third friend, Andra. Mary is in a good mood. Then, when Andra enters the room where Mary is chatting with Lucas, Mary instantly and unexpectedly notices in Andra's facial expression that she feels miserable. In a wink her mood, expression and attitude change and this suddenly emerging aspect of her friend's condition affects her subsequent activity. From one readiness-to-act (continuing the nice conversation with a friend) Mary shifts to another

(comforting Andra, inquiring what is wrong, etcetera). (Klaassen et al 2010, 56)

Notice that, as set up, the case itself does not refer to social norms, or to the encounter being embedded in any particular structure or collection of structures. That is a positive thing. But there are a number of other features of the description that do make the case potentially problematic (as a generalisation): first, Mary is said to 'instantly' notice Andra's expression, and we may wish to query why it is necessary to limit Mary's unreflective, non-deliberative abilities to an immediate recognition; second, Mary is said to notice 'that' Andra 'feels miserable', thereby suggesting a) that the mode of Mary's noticing is a 'that' claim (which presumably Mary would be prepared to back up), and b) that Andra really does feel miserable and that what makes Mary's action so impressive is that she recognises this, or has such a sensitivity as to be able to detect this; and third, that once Mary notices how Andra feels she knows instantly what to do, as if she was guided by a script associated with noticing that Andra feels miserable.

An alternative description of the case would depict Mary appreciating (not necessarily immediately, but over a certain period of time), from how Andra appears to her, Andra's misery, and then orienting herself to engaging Andra in a way that Mary feels may alleviate what Mary perceives as Andra's misery. Putting matters this way is more careful, for it allows us to see that Mary may not be recognising some fact in the situation, but rather relating tentatively to how Andra might be feeling, and it also allows us to see that she may proceed with care (not absolute confidence) about what to do in reply to how she relates herself towards how Andra appears to her. In other words, without setting up the situation as one where we (as the theorists) assume that Andra is really miserable and thus describe Mary as having abilities to detect certain properties (according to which Mary can justify her judgement that Andra is miserable), we leave open the possibility of the description of a more tentative, explorative ability to appreciate how things appear to one and what one ought to do.

How do – on the back of their own description – Klaassen et al explain the case? They do so separately in two different passages in the paper. In the first passage, Klaassen et al say that:

Mary immediately and with instinctive fluency responds appropriately to Andra's distress (in a quite complex social context). Just as Wittgenstein's

expert craftsmen are absorbed in their work, Mary is absorbed in her social engagements. Her action can be seen as an example of social instinctive normative action, because from a first-person perspective Mary's response to Andra is an instance of 'blindly' acting in accordance with rules of social interaction. (Klaassen et al 2010, 56)

Notice that the explanation makes it clear that the way Klaassen et al imagine the case is that there is something in the situation that it is appropriate to do (comfort Andra). This way of imagining the case is the result of two things: first, that it is assumed that Andra really is miserable; and second, that it is thought that the only way Mary's action could be appropriate is if there were 'rules of social interaction' in place that Mary's action could be described as in accordance with. The problem with this way of explaining the case is that Mary is positioned passively with respect to what can become normatively relevant, and, as a result, her ability to interact with Andra is also described passively (as merely recognitional – she merely recognises that Andra is miserable, and that this means she ought to comfort her). What this misses is that Mary may not be recognising any properties, but rather be tentatively appreciating, in a way perhaps unfamiliar to Mary, how Andra's appearing to her suggests to her that she ought to do something.[201] Further, Mary may interact with Andra in a way that we (as observers of the scene) might come to appreciate us appropriate, but not in a way that is familiar to us, e.g. imagine that, instead of comforting her (asking what is wrong etc) Mary tells Andra a joke, and this puts Andra into a better mood. The point is, if we impose on the action an explanatory scheme whereby there exists a social norm that when one detects that someone is miserable one ought to comfort them, we also miss the sense in which Mary may do something else that we may also come, eventually, to think could be appropriate (of course, we may also not: the point is to leave room for this indeterminacy about how something might come to be normatively appreciated).

The point here is to allow for Mary being more active vis-à-vis the appropriate thing to do: Mary can surprise us (indeed, she can also surprise Andra who may have been expecting being comforted, but in fact liked that Mary instead offered a joke). Allowing for Mary to be more active vis-à-vis the normative means that it is harder to say whether Mary does something appropriate or not, but that is as things should be:

---

[201] Mary need not be recognising 'what' she ought to do; indeed, 'what' she does may be less important than Andra perceiving that Mary is paying careful attention to her, expressing concern in how Andra is feeling – and thus not presuming to know what Andra is going through.

Mary does not know – she cannot guarantee – that how Mary appreciates Andra will not be challenged by Andra (in saying this we need not assume that Andra knows how she feels – she too may be tentatively relating with herself, i.e. there is no first-person authority assumed here), and she can also not guarantee that the way she sees herself as responding to Andra (telling a joke) will be appreciated as appropriate by Andra (perhaps Andra will feel hurt that she was not comforted, but perhaps she will not; perhaps Andra will not even think that what Mary does is 'telling a joke', but instead that Mary is 'ignoring' or 'trying to distract away from' what Andra thinks Mary must see as Andra's misery). Let us be clear that the point here is not that perhaps what ought to happen is that Mary ought to tell a joke; we are not suggesting an alternative explanatory scheme ought to apply; what we are trying to do is to open up the description and explanation of the case (precisely by unravelling any imposition of what it would be appropriate to do) in a way that allows us to see that Mary's abilities may need to be described in a more active and creative, and thus also tentative, fashion.

There is another aspect of Klaassen et al's explanation in the above passage we ought to highlight, namely that although Rietveld et al want to extend their explanation (of how we can act adequately, yet unreflectively) to everyday action, and thus go beyond discussing experts engaging with familiar material, they end up presenting everyday life as precisely a context of rules to be mastered in the same way that (on their view) an architect or tailor allegedly masters their context, i.e. they internalise the rules so much that they can instinctively, or blindly, act in accordance with them. What Klaassen et al do, then, is present social life as a domain of rules that we can all master to such an extent that we become social experts (thereby being able to act socially appropriately, yet unreflectively).

It has already been argued earlier in this thesis (in Chapter 13) that there is another way of understanding expertise, i.e. as an ability not to act more instinctively and immediately with the rules allegedly governing the domain (and determining what is appropriate) but rather slowly and with difficulty creating new ways of doing things that may establish themselves as ways others might wish to emulate (of course, they may also not establish themselves that way – this is left open). Similarly (and perhaps even more so) with social life: it is possible that the more experienced we get the more we appreciate the diverse ways in which we can experience the lives of others (how different they can be, how they can care about different things, how they

can surprise us), such that we become more capable of being active, and not passive, vis-à-vis the normative (creating new ways of relating normatively with others). Memorable examples of this abound in the history of literature and cinema: consider, for instance, *The Elephant Man* (1980), where someone was able to show courage in how they related normatively with someone in distinction from how others related with them. The point here is not a moralistic one: it is not to preach and say that the Elephant Man ought to be related with in a certain way; it is instead to leave open the description of the case that allows for persons being more active vis-à-vis the normative (vis-à-vis what they and others around them may come to experience as appropriate).

We mentioned earlier that there was a second passage where Klaassen et al explain the above example. This explanation refers to certain concepts introduced earlier in the paper, which we will need to return to, but let us first quote the passage:

> …it can be argued that when investigated on a micro-timescale, it is plausible that Mary's first response to Andra's expression is non-conscious or pre-reflective. Just as in the example of the Brazilian tailor discussed above, Mary is probably affected by the situation before she has become reflectively conscious of how Andra feels. That this is a case of motion-affect is clear if we focus on, for instance, the changes in her facial muscles, which underlie Mary's change in expression, or on her initiation of a comforting action… With regard to instinctive normative action the hypothesis is that the individual's embodied responsiveness to possibilities for action changes unreflectively and that the two relevant stages in this process are the emergence of valence and directed discontent, respectively. When Mary perceives Andra's expression, the development of valence already implies anticipation on several appropriate possible actions, whereas the route to other (inappropriate) possibilities for action is cut off. (And what is more, these inappropriate ways of action do not attract the skilful individual). (Klaassen et al 2010, 66)

The Brazilian tailor referred in the above passage is described as someone who, when faced with a certain fabric, immediately 'sees that the pattern of the fabric he just received is incorrect' (Klaassen et al 2010, 63). The tailor's feeling is broken down into valence and directed discontent. Valence (a concept that Klaassen et al borrow from Varela and Depraz 2005) is said to refer 'to the ability of the individual to determine whether he or she should move towards or away from something that appears in his or her environment', the point being that this ability is at once a kind of 'motion-affect', i.e. 'a kind of tension that can be characterised by both the (pre-

reflective) experience of attraction/repulsion and the related behavioural tendency to move towards/away from the triggering features of the situation' (Klaassen et al 2010, 64). The immediate perception of objects, then, is correlated with certain possibilities of action, and the way that Klaassen et al explain this is that 'meaningful objects' are 'value-laden', such that they are associated for us (presumably in our bodies, and / or in the ways our brains work) with certain 'repertoires of possible actions' (Klaassen et al 2010, 64). Put this way, it is not easy to see what directed discontent captures that valence does not already include, and indeed Klaassen et al say that 'valence resembles directed discontent' (Klaassen et al 2010, 65), but they distinguish it in the end on the basis that directed discontent is a more mature affect whereas valence is more immediate and more reflexive (appears on a shorter time-scale). Directed discontent (which we looked at in detail in Chapter 13) is a further process that may grow on top of, or follow, valence, but it is really valence that is doing the crucial work in the explanation (as it is in the above explanation of Mary's case). Here is how Klaassen et al further elaborate on the importance and character of valence:

> It is crucial that the appreciation of an object or event (as being correct or incorrect, satisfactory or unsatisfactory, etc.) is already being formed on the micro-timescale of valence. The architect's event-related valence will be very different when he sees an incorrect door (in context) than when he sees a door that is exactly right in its context. His responsive is perceivable to others. In the situations of the craftsmen at work, social norms concerning appropriate action are already being reconfirmed (or not)[202] at the very moment at which we speak of valence. Therefore we wish to suggest that valence, as a precursor to directed discontent, can be seen as the most basic moment at which situated normativity becomes perceptible. Valence is in these cases of instinctive normative action not some marginal fluctuation but the earliest integrated tendency of an individual who is developing an adequate response to the situation. In other words, we can understand the person's experience and expression of (directed discontent-preceding) valence as the most basic form of instinctive normative action. (Klaassen et al 2010, 66)

What we see here is that valence is made subordinate to a particular scheme that determines what is appropriate – in this case, a way in which what is appropriate is established by the social norms governing the correctness of fabrics. This is then extended, by analogy, to Mary's case. In this way, we encounter the same problem: Mary's abilities are described too passively as a result of her reaction being seen to be

---

[202] By 'not' here Klaassen et al appear to mean that when the norms are not followed, and thus violated, they are not reconfirmed.

effectively determined by what is imposed as the appropriate thing to do in accordance with the scheme of appropriateness that is assumed to apply. We do not see the tailor being able to change what he and other persons in that domain might come to appreciate as appropriate: instead, what happens is that as a result of years of training the tailor (and Mary) is able to immediately and unreflectively – with ease and fluency – to recognise what is appropriate.

It should be emphasised here that there are many aspects of Klaassen et al's discussion that we can build on, and that need not be associated with the governance view. Consider, for instance, their reference to the following quote from Wittgenstein:

> How do I find the 'right' word? How do I choose among words? Without doubt it is sometimes as if I were comparing them by fine differences of smell: *That* is too…., *that* is too…, – *this* is the right one. – But I do not always have to make judgements, give explanations; often I might only say: 'It simply isn't right yet'. I am dissatisfied, I go on looking. At last a word comes: '*That's* it!' *Sometimes* I can say why. This is simply what searching, this is what finding, is like here. (Wittgenstein 1953, 218; quoted in Klaassen et al 2010, 55-6; original emphasis)

What is interesting in this process is that it allows for flexibility as to what may be the 'right word', i.e. Wittgenstein explains a process of considering alternatives, the point being that the outcome is not determined (and not described as being determined). The outcome, it should be added, can still be one where we have the experience, at the end, that that is the right word, but the important thing is that the case is not described in a way where our expertise depends on us detecting just that word (which is said to be 'right' on the basis of some other criteria). In other words, having gone through the process of searching we may end up with something that strikes as right, correct or appropriate, and we may then be tempted to think that this was the answer that was waiting for us all along. It is important to see that we can have such an experience, but it is also vital that we do not take it too seriously, i.e. that we do not think that this experience suggests that there was just that word waiting for us to be recognised as the right, correct or appropriate one.

The point here is not that appropriateness is determined by our say-so; on the contrary, the point is to say that the case needs to be described in a way where appropriateness is not determined, for this leaves the process open (allowing it to be possible for us to treat the external constraint as unchangeable and be guided by it, but

also allowing us to be active and create what it might be relevant to consider in appreciating some new possibility as an appropriate one).[203]

Notice that if we are to make room for the latter possibility, we may also need to distance ourselves from the idea – which is there under the surface in Klaassen et al – that 'unreflective' or non-deliberative means immediate, reflexive, instinctive. In other words, the process we may need to make room for may need to be both unreflective and protracted, i.e. explorative rather than reflexive, and hesitant rather than immediate. To act unreflectively, then, need not be a matter of acting fluently, easily, or confidently; it may be groping, orienting, and tentative, and still be highly pragmatic, embodied, non-deliberative.

Notice, too, that in Wittgenstein's description, he explicitly says that the process need not be one that involves making judgements or giving explanations. This suggests the possibility of a process of appreciation – but, again, not an immediate one destined to recognise something in particular, but one that explores, searches, moves through possibilities (that it itself creates as possibilities; it is vital not to describe the alternatives as pre-existing the process of searching for them), and (potentially, though not necessarily) settles finally on something (without that something being describable as that which the person was always going to find).

The ideas we see in Klaassen et al can be found in many other theories, and it is time now that we consider these. Although all are different in details, they also all have in common the description of persons' skills, abilities, dispositions (whatever they are called) as being necessarily sensitive to necessarily shared determinants of appropriateness and inappropriateness.

B. The Background á la Searle

One of the most prominent references to the idea of the background can be found in John Searle's account of social life (it is no accident that Searle and Dreyfus both refer to the background, both having been colleagues at the same institution for a long time). Arguably, Searle first introduced the notion of the background into his own work in order to 1) avoid claims that he was over-intellectualising the mind; and 2) not 'run afoul of the Wittgensteinian idea that rules (and other forms of intentionality)

---

[203] The focus in this paragraph is on a person-to-environment relation, but the matter can be re-phrased for a person-to-person relation, where the point is that what might be treated as relevant and important can be created for one by the other.

are not self-interpreting but rather determine conditions of satisfaction only against a background of nonrepresentational mental phenomena' (Hershfield 2005, 276). It was, in particular, this second motivation – to not keep presupposing (and thus avoid triggering the ensuing regress if one presupposes) an intention behind an intention, or a rule behind a rule – that Searle's background is composed of non-intentional skills and dispositions that exist in the form of causally efficacious neural pathways.

It should immediately be pointed out, and indeed stressed, that Searle's account of the background is designed to avoid the postulation of 'unconscious rule following.' Searle wants to avoid this for he does not want to make it the case that one would need to first consciously learn the rules, and then come to internalise them such that one could be said to be following them unconsciously. Instead, what he argues for is that 'we should say, first (the causal level), the person behaves the way he does, because the has a structure that disposes him to behave that way; and second (the functional level), he has come to be disposed to behave that way because that's the way that conforms to the rules of the institution' (Searle 1995, 144).

We shall examine this position more in a moment, but first we must note that Searle puts himself in this position because of the priority he assigns to 'the rules of the institution' or, in any other words, to the rule-constituted institutional reality of social life. This is not the occasion for a full-scale review of Searle's social metaphysics, but it is important to note a few of its features (we already noted a few of them in Chapter 7). There are four elements of Searle's approach to social reality: the assignment of function; collective intentionality; constitutive rules; and the background. It is the first three that we have to briefly define here before discussing the fourth.

What does Searle mean by the assignment of function? According to Searle, persons, and some animals, impose functions on objects – these being either 'naturally occurring objects' or 'those created especially to perform the assigned function' (Searle 1995, 14). There are various kinds of functions we can assign – functions that refer to uses we make of objects (agentive functions); functions that refer to naturally occurring causal processes to which we assign a purpose (non-agentive functions); and functions that enable marks to symbolise, represent or stand for certain other things (meaning functions, let us call them). What is common and critical about them for Searle is that such functions may be assigned by us (they are observer-relative after all) but not mediated by our awareness that we are so assigning

them: 'functions may be imposed quite unconsciously, and the functions once imposed are often – so to speak – invisible' (Searle 1995, 22).

By collective intentionality Searle means not only (and not primarily) that persons can co-operate with each other, but that 'they share intentional states such as beliefs, desires and intentions' (Searle 1995, 23), e.g. 'if I am an offensive linesman playing in a football game, I might be blocking the defensive end, but I am blocking only as part of *our* executing a pass play' (Searle 1995, 23; original emphasis). In this example, I can have an intention to execute a block, 'but I have that intention only as part of our collective intention to execute a pass play' – hence the idea that the 'crucial element in collective intentionality is a sense of doing (wanting, believing, etc) something together' such that any 'individual intentionality that each person has is derived from the collective intentionality that they share' (Searle 1995, 24-5). Such collective intentionality is signalled and can be expressed by the form 'we intend' (and not the addition of one 'I intend' with another 'I intend'). Searle refers to 'any fact involving collective intentionality' as a 'social fact', a sub-class of which is 'institutional facts', such as 'the fact that this piece of paper is a twenty dollar bill' (Searle 1995, 26).

What about constitutive rules? This is an element that is familiar from our previous discussions (especially in Part II of this thesis). Briefly, on Searle's account, constitutive rules have the form: 'X counts as Y' or 'X counts as Y in context C', with the familiar examples including the rules of chess being constitutive of chess 'in the sense that playing chess is constituted in part by acting in accord with the rules' (Searle 1995, 28; what is meant by 'in part' is not elaborated upon), and being extended to the general social domain in that it is by (and only by) virtue of constitutive rules that certain facts are institutional facts (pieces of paper count as money only by virtue of certain constitutive rules). These constitutive rules, it should be added, are distinguished from regulative rules, which are rules that apply to previously existing activities: e.g. 'the rule "drive on the right-hand side of the road" regulates driving; but driving can exist prior to the existence of that rule' (Searle 1995, 27).

In his *The Construction of Social Reality* (1995), from which the above account is taken, and in many works before and after then (most recently in *Making the Social World*, 2010), Searle has developed the above in great detail, answering many objections along the way. It is outside the scope of this thesis to examine

them.[204] Our interest, as noted above, is in the fourth element of Searle's account, namely his reference to the background. Searle examines this element in detail in chapter 6 of *The Construction of Social Reality*, which is entitled 'Background Abilities and the Explanation of Social Phenomena.' The dilemma that Searle faces – having begun with the first three elements, and especially given his employment of the constitutive strategy – is how to make his metaphysics psychologically and phenomenologically plausible. This dilemma is clearly articulated in the opening paragraph of that sixth chapter:

> I have said that the structure of human institutions is a structure of constitutive rules. I have also said that people who are participating in the institutions are not conscious of these rules; often they even have false beliefs about the nature of the institution, and even the very people who created the institution may be unaware of its structure. But this combination of claims poses a serious question for us: Under these conditions, what causal role can such rules possibly play in the actual behaviour of those who are participating in the institutions? If the people are who are participating in the institution are not conscious of the rules and do not appear to be trying to follow them, either consciously or unconsciously, and if indeed the very people who created or participated in the evolution of the institution may themselves have been totally ignorant of the system of rules, then what causal role could the rules play? (Searle 1995, 127-8)

To this, we should add, that the rules in question 'are in general not codified', or when they are, 'most of us are unaware of' them, and even 'if we were aware', we 'have to know how to interpret and apply the codified rules' (Searle 1995, 128).

We have already mentioned that Searle notes, but dismisses, a reply based on the idea of unconscious rule-following, and offers instead his notion of background. Here, we are not so much interested in his criticism of the unconscious rule-following strategy, but in his account of the background.

The background is composed of 'nonintentional or preintentional capacities that enable intentional states of function' (Searle 1995, 129). 'Capacities' is an umbrella term – which may also include abilities, dispositions, and tendencies – that points to structures existing at the neurophysiological level and thus to the causal capacities of the brain (it is at this causal level that we should also understand the sense of 'enablement' in the above definition; Searle 1995, 129-30).

---

[204] A recent and very instructive review of Searle 2010 is Roth 2012.

Searle identifies numerous functions of the background: for example, that it enables linguistic interpretation to take place; that it enables perceptual interpretation to take place; that it structures consciousness; that it facilitates certain kinds of action readiness; and that it disposes persons to certain sorts of behaviour (Searle 1995, 132-7), but what is of particular interest to us is not what distinguishes these functions, but what holds them in common, i.e. how the background is said to work in general.

Under most of the above functions identified for the background, Searle provides different examples; what is interesting for us is that in each of them, the role of the background is to block alternatives and set the person on the required path (as we shall see – although it might already be obvious – the required path is required, and determined to be required, by the rules). Thus, for instance, although there is (or, better, precisely because there is) 'radical undetermination' in what 'is said by the literal meaning of a sentence', we need something (or so Searle seems to assume) that will block the 'indefinite range of ridiculous but still literal interpretations of this or any other sentence' (Searle 1995, 131) – this 'something' is precisely the background: 'the only thing that blocks those interpretations is not the semantic content but simply the fact that you have a certain sort of knowledge about how the world works, you have a certain set of abilities for coping with the world, and those abilities are not and could not be included as part of the literal meaning of the sentence' (Searle 1995, 131). Similarly with perceiving things: 'we are able to see the figure as either a duck or a rabbit [notice that these are the only two possibilities], because we bring to bear on the raw perceptual stimulus a set of Background skills; in this case we bring the ability to apply certain categories' (Searle 1995, 133). And again with my finding 'certain kinds of jokes and not others' funny (Searle 1995, 136), or with my expectations and forms of action readiness: in all these cases, alternatives are blocked and a certain range of what I can experience is said to inhere my neurophysiological make-up.

The question to ask at this point is as follows: why does Searle require it to be the case that alternatives be blocked? Once we pose this question we begin to see a little clearer the way the argument works. The question is how much the background secures: does it secure that I will, without fail, laugh at those jokes and only those jokes, or that I will interpret the sentence in that way and only that way – or does it leave open the possibility that I may laugh at a different kind of joke (precisely not a 'recognisable kind', whether for me or even perhaps for anyone else) or that I may

interpret the sentence in a way it has not been (including by me and / or others) previously interpreted? If it does leave open the possibility, how does it explain it?

There are two moments in the chapter when Searle allows for a person to be surprised or to depart from the background: first, he points to the ability of surrealistic painters to break with the certain ways of representing persons ('the three headed woman'; Searle 1995, 134); and second, he points to how he 'would be absolutely astounded if a skier suddenly came through, or if an elephant simply walked into the room' (this is in the context of a lecture hall; Searle 1995, 136). Searle does not explain this capacity for surprise or change, except by saying, with respect to the second, that 'such things could happen, but they definitely are not the sort of thing that the Background makes me ready for' (Searle 1995, 136). Now if the background does not make me ready for them, but they can happen, then this suggests the background has its limits: it explains that which it predicts will happen, but it does not explain that which it does not predict (of course, once something happens, the background might be said to latch on and then be able to explain it when re-occurs, but presumably there is a first time something happens – e.g. a first time a Surrealist painting is painted). But how to explain that which the background does not predict / does not prepare us for? The question may strike one as somewhat paradoxical in the context of Searle's system, and it is because Searle's cases are described in such a way where we do not need to ask the question. Put differently, the way Searle sets up the background is misleading, for it is based on two mistaken assumptions: first, in thinking that alternatives must be blocked, whereas in fact they might be (actively treated by us as unblockable) but also might not be (there is no determinacy here); and second, in thinking that there is something that can block them (saying something blocks them has the effect of making us the passive re-enactors of our past).

All this becomes clearer when Searle – unsurprisingly, given the usual association with the constitutive strategy – appeals to how the background works in the contexts of games. Here is an instructive passage:

> The key to understanding the causal relations between the structure of the Background and the structure of social institutions is to see that the Background can be causally sensitive to the specific forms of the constitutive rules of the institutions without actually containing any beliefs or desires or representations of those rules. To see this, let us start with a simple example. Suppose a baseball player learns a set of rules, principles, and strategies. But after he gets skilled, his behaviour becomes much more fluent, much more

melodic, much more responsive to the demands of the situation. In such a case, it seems to me, he is not applying the rules more skilfully; rather, he has acquired a set of dispositions or skills to respond appropriately, where the appropriateness is actually determined by the structure of the rules, strategies and principles of baseball. The basic idea…is that one can develop, one can evolve, a set of abilities that are, so to speak, functionally equivalent to the system of rules, without actually containing any representations or internalisations of those rules. (Searle 1994, 142)

The key part of this passage is the sentence that says that appropriateness is determined by the rules. The point is that it is only as a result of this assumption that Searle can make it look as if the background works as it does, for it is this assumption that does the work of blocking the alternatives – of making it seem as if I could not do anything but act in accord with the rule. The way the game image works in Searle is that there is a structure of possible actions; certainly, the background might be said to prepare us for this structure, but only if we define that structure so tightly that no other action is possible in it; this is, in effect, what the constitutive strategy does, for it says that we cannot play baseball (perform the actions required by baseball rules) if we do not conform to the rules. This is what makes it difficult to see the limits of the notion of the background. Once we see that appropriateness is not determined by the rules, and that it is possible act (to play the game) in a way that is not either in accord or in violation of the rules (here 'the game' must be understood to be what we and others might appreciate as still playing baseball), then we also see that the background works to explain only that which we expect will happen (based on treating the rules as unchangeable and determinative of what we could do).

On Searle's view, our background dispositions, abilities, skills, tendencies, and so on, are all sensitive – or better, happen to be sensitive – to 'the specific content' of the rules (Searle 1995, 142) or 'to the rule structure of the institution' (Searle 1995, 144). The above baseball example can be a little misleading, for it might lead people to think that Searle is still requiring persons to first learn the rules consciously, but in fact Searle allows for the person to simply 'grow up playing baseball', where baseball is defined to be playing in accordance with the rules of baseball (see Searle 1995, 144-5).[205] Thus, when extending this to social life, Searle does not argue that we 'the man who is *chez lui* in the social institutions of his

---

[205] This might remind us of Chomsky's idea (see Section 8.C.) of 'rules growing in the mind' (except that here, on Searle's view, they 'grow in' one's neurophysiological make-up).

society, is at home because he has mastered the rules of the society, but rather that the man has developed a set of capacities and abilities that render him at home in the society; and he has developed those abilities because those are the rules of his society' (Searle 1995, 147). Our answer to this is simple: we are not always at home in our society, and our society is not such as to guarantee that we will always feel at home (we could say: it is not the sort of place where one must feel at home or that is a home with set paths for doing things). The rules of baseball, or anything we might refer to as the rules of society (some set of normative resources, say, that we may contingently find useful to appeal to), do not control what could happen or how we might act or interact, or indeed how we might appreciate something as good or bad. It is not a foresworn conclusion that we will always play the game of baseball in that way, always laugh at the same kind of jokes, or always interpret words in the same way – the trouble with the background, and the accompanying examples, is that it makes it seem that way (and then describes us as being prepared, happily, to keep repeating and reinforcing those same ways).

All this is tied very closely in Searle to the appeal to a 'normative dimension'. This becomes very clear in the following passage:

> Somebody might say, 'Why do you have these rules at all? Why don't you just have some kind of behaviourism? These things just happen, people just do these things.' The answer is that where human institutions are concerned, we accept a socially created normative component. We accept that there is something wrong with the person who when the baseball is pitched at him simply eats it; something wrong with the person who doesn't recognise any reason to do something after he has made a promise to do it; something wrong with the person who goes around spouting ungrammatical sentences. And all of these cases involve something wrong in a way that is different from the way there is something wrong with the man who stumbles when he walks; that is, there is a socially created normative component in the institutional structure, and this is accounted for only by the fact that the institutional structure is a structure of rules, and the actual rules that we specify in describing the institution will determine those aspects under which the system is normative. (Searle 1995, 146-7)

What we find here is that how we conceive of the normative dimension matters enormously for how we conceive of the abilities of persons. If we think that things could not be appropriate (or inappropriate) without some set of rules determining it, then we may go on, as Searle does, to depict persons' abilities as ones that prepare them to act precisely in those ways and not others (without them realising that that is

what they are doing). Searle's view is an example of how we can come to think that what we 'can' (in the sense of what we are able to do) do is controlled by what we 'should' do (on the theorist's account of what ought to be done when, e.g. playing a game), but we can unravel this view, and thus see that what we 'can' do is much broader, when we also see that what we 'should' do is broader too, i.e. when we see that what we 'should' do is not determined by any set of rules (even if it may sometimes be helpful to appeal to those rules in guiding ourselves and evaluating ourselves and others).

At one point in his chapter on the background, Searle speaks of there being 'ungrounded way[s] of acting', by which he means acting in accordance with a rule but not with the intention to act in accordance with it (Searle 1995, 140), but what he fails to consider is how we can act in a way that is not necessarily in accord with a rule. It is because, in other words, Searle sets things up in such a way that he thinks we must act in accordance with a rule that he employs the background to explain how we come to act that way (i.e. how we come to be able to always in accordance with a rule). What we have been arguing is that we see how limited the notion of the background is as soon as we see that there is no 'must' in the sense of necessity there: we could, but we might not, be taken to act in accordance with a rule (if others think the rule is relevant and applicable to evaluating us); we can do things in a way that is irrelevant to the rule's applicability, and nothing in the background (conceived of as sensitivity to the rules) prepares us for this. Appreciating the potential: 1) irrelevance of a rule; 2) creation of new rules; 3) for coming to care about things in a way we have not cared about previously; and 4) for introducing (perhaps by accident, and without realising it) ways of playing games – all these, and more, are examples of things we do that contribute to and are part of a normative dimension, just as long as we do not define, from the beginning, that normative dimension in such a narrow way as to require there to be determinants of what is appropriate and inappropriate.

Searle seems to want to say that whatever does happen was always being prepared for, e.g. that if we laugh at such-and-such a joke, it is because that joke really is part of the category we always laughed at (we were just mistaken in what category we found funny – in other words, that we were always destined to find that funny, but we just did not realise we were). The same form of argument appears (at first blush) more persuasive in the context of what is appropriate and inappropriate (and that is precisely with the normative dimension is appealed to), i.e. it looks like

(when we describe the cases in a certain way) that what we find appropriate or inappropriate could not but have been found appropriate or inappropriate in the way we found it to be so. But to say this is to be put the cart before the horse: what I will find funny is not determined either by what I have found funny in the past or by any conception of what I have found funny in the past – I may find a new joke funny in a way I have not found anything funny before; similarly, when we are faced with a new sentence, we may find it ungrammatical because we evaluate it as sufficiently alike other things we have found ungrammatical and because we have found those things ungrammatical under a rule, and as a result say that this new sentence is also ungrammatical under that rule – but we may also find that sentence sufficiently unlike anything else, and thus either grammatical or ungrammatical, but grammatical or ungrammatical not under any pre-existing rule (one could say here that the relevance of the rule is not determined by our ability to apply the rule).[206] Put differently, just because some way of doing things – some way of playing baseball, some way of finding things funny, some way of finding sentences grammatical – presents itself as (because we can, in some moment, actively treat it as) relevant, does not mean that all we can do is defined and characterisable as either in accordance or in violation of that relevance. What we need to avoid, when offering cases of the normative, are descriptions that make it seem as if we could not have done anything but what we happened to do, such that the status of what we did – whether it was appropriate or not – was always and already predicted / prefigured in something prior and external to our action.

In case this is not yet clear, let us try to articulate it once more. Take the ability to play football (soccer in America). If one pictures football as requiring certain kinds of actions, and one defines this tightly, such that the actions are defined by rules (that may or may not be known to the participants) and the rules determine what is both possible (what can be done) and appropriate (indeed possible because appropriate), then one is going to force the description of the abilities (and their

---

[206] A curious recent example we could mention here is the inclusion of 'words' such as LOL and OMG in the Oxford English Dictionary (OED). As originally reported on March 25 2011 by the Associated Press (see Freeman 2011), there has been passionate debate amongst both concerned citizens and grammarians about the 'appropriateness' of this inclusion, and thus of the relevance of the criteria for the OED as to what ought to count as a 'word'. The whole point here is that we could not have predicted the emergence of such 'words' as LOL and OMG – although this already puts it too strongly, for part of the point is that it continues to be an issue as to whether LOL and OMG ought to be treated by us as 'words'. The 'ought space' is what enables debate and disagreement, and not that which secures agreement in advance.

development) to be a matter of coming to be able to and repeating on any occasion those very actions. But if one does not picture football in this way, and notices that there is room within football to perform different kinds of actions, and that what it is possible and appropriate to do within football is not fixed, then one also allows for different persons to develop abilities that are not necessarily describable as abilities to re-enact or repeat the same actions (the ones that also happen to be in accord with, i.e. required by, the rules).

There are a number of ways in which one can picture football in this different way. Consider, for instance, styles of playing football, and notice that one can particularise this at various levels: e.g. different styles of playing football (and thus also different abilities to develop) in Brazil rather than Germany, and between such-and-such a club within Germany, and between such-and-such a generation within that club in Germany, and so on. The point is that if one begins with football defined by a certain set of rules, one will not see how the abilities players in those different groupings may be different, and different in ways that are not explicable on the basis of the rules of football.

But one can go further than this: is it the case that the rules of football are so fixed, and specify the required actions to such an extent, that all of us who play football develop abilities to perform those very actions and those very actions happen to be in accordance with the rules? Consider the phenomenon of tackling. Players in bygone times may have developed abilities that enabled them to be quite rough with each other in the penalty box, but as times have gone on, players have generally had to develop abilities where they could, say, defend effectively without rough tackles. The point here is that at no point might there even have been any rules, and any attempt to articulate rules would have been, and still today would be, very controversial: what sorts of tackles exactly are not allowed, and which are permitted? Can rules define – specify – the exact kind of roughness permitted and the exact kind not allowed?[207] Or take the phenomenon of diving (looking for a foul): at a certain

---

[207] The point, of course, is that they cannot: instead, at best, any rules that there were would both require and enable an evaluation to take place, not in the sense in which they determine it, but precisely in the sense that they do not, i.e. a person needs to exercise an indeterminate evaluation as to the sense in which the rule may be said to have been followed or violated (more commonly, violated). The evaluation is indeterminate because there is nothing either in the rule or in us (in our experiences, say, of observing football) that will determine whether we appreciate a certain tackle as in conformity or in violation of a rule. Of course, there need not be rules to enable us to evaluate tackles as appropriate or not.

point in the game (and perhaps even now to some extent) persons were not quite clear about whether, and if so how, such practices should be banned (because it is very difficult to tell whether a player is diving, or has just lost balance say); one could say that some players developed abilities to dive (masking the dive well), but this ability is not explicable on the basis of developing an ability to repeat an action that is in accord with a rule.

The discussion above might remind us of our previous discussion of games (in Chapter 7), and how one can simply not understand them adequately or exhaustively on the basis of any rules that we might, at any particular time, associate with them. What Searle does is that he keeps the rules in place, thinks that the rules specify the required (possible) actions, and then describes persons as learning and repeating those actions (these actions then being able to be said to be sensitive to the rule-structure of the game), but as we have argued above, the idea that games (or any activities, let alone social life generally) is a rule structure, where the rules specify exactly what can (because it ought) to be done is simply implausible: it fixes, and over rigidifies, games, and as a result narrows our picture of the development and abilities of persons.

One final thing in this respect: notice that the same problems arise whether one makes the further move to say that the actions are required by the rules, or whether one simply says that the game (or activity) is patterned, i.e. that the activity is constituted by certain specified ways of doing things (certain actions that are repeated): this too will fail because no such specification is going to be exhaustive (it is at best a rough guide to how to play the game at a certain time and place, given by someone who understands the game in a certain way). The point here is that in both cases we begin with a certain way we understand some activity to be ordered – either by rules specifying actions, or by actions as patterns – and we then describe persons as learning to and participating in an activity understood in this way; and what we fail to see is that although we can give a more or less accurate description (descriptions that would be more or less successful as guides to how to play football), this description is always limited, and will not pick up on the way the activity always remains somewhat open, both in terms of diverse styles or regions of the game that are uncertain, and in terms of the fact that it is simply impossible to specify actions required or reproduced so tightly (activities are simply not the sorts of things that can be so tightly specified); and the point is that when we fail to see this, we also narrow down our picture of persons learning and acting (or interacting).

At one point in his chapter Searle says that his account is congenial to the notion of the habitus developed by Bourdieu (he says that it 'is about the same sort of phenomena that I call the Background': Searle 1995, 132). Let us, then, turn to briefly examine the habitus and see how it depicts the capacity of persons to act unreflectively, yet appropriate socially.

C. The Habitus

Bourdieu is very aware of the dangers in 'sliding from the model of reality to the reality of the model' (Bourdieu 1990, 39). However, like Searle, he confines this mistake to what he calls 'legalism', i.e. the idea that one can 'slip from regularity, i.e. from what recurs with a certain statistically measurable frequency and from the formula which describes it, to a consciously laid down and consciously respected ruling…or to unconscious regulating by a mysterious cerebral mechanism' (Bourdieu 1990, 39). In other words, Bourdieu wants to keep the sense in which social life is ordered, and wants to picture persons learning to participate and participating in such orders, but he does not want to: 1) assume or posit that those orders were imposed by some authority in the form of 'consciously devised and sanctioned rules' (Bourdieu 1990, 39); or 2) assume or posit that persons must either consciously or unconsciously learn those rules, and then consciously or unconsciously process them. But notice that Bourdieu does want to keep the sense in which there are patterns of acting and interacting, and there are 'formulas' that can describe these patterns, such that persons can be described as learning to and reproducing the patterns.

Here we can ask: why is this not another version of the same thing Bourdieu wants to avoid? Cannot persons be said, on this view, to consciously or unconsciously learn to, and consciously or unconsciously process, the 'formulas' (which happen to correspond to the patterns that constitute the order in which persons are pictured as participating)? Of course, one can point to differences between formulas and rules – for instance, no one would want to claim, as they might be tempted to in the case of rules, that some authority first devised the formulas and then others come to conform to them – and these are important differences – but for our purposes, the general strategy is the same: whether said to be planned or not, social life is imagined as ordered, and ordered in such a way that there are patterns of actions and interactions that are reproduced, such that persons are described as learning to reproduce and thereafter regularly reproducing those patterns. As we have argued in the case of

Searle, this works only from a distance, i.e. it works when one does not challenge the description of the patterns (the very fact that we can challenge the descriptions shows that they must be a limited way of understanding the activity).

Now this will seem a very unfair criticism of Bourdieu, for Bourdieu is keenly aware of, and eager to avoid, what he calls a kind of 'objectivism idealism' according to which 'the social world is seen as a representation…or a performance…and practices are seen as no more than the acting-out of roles, the playing of scores or the implementation of plans' (Bourdieu 1990, 52). Bourdieu does not want to depict the 'objects of knowledge' as merely 'passively recorded', and he insists on the 'active aspect of the apprehension of the world' (Bourdieu 1990, 52). But here we must look very carefully at just what Bourdieu means by 'active', about the 'structuring' aspect of those 'structured' dispositions. Before we do that we must allow Bourdieu to define what he means by a habitus:

> The conditionings associated with a particular class of conditions of existence produce habitus, systems of durable, transposable dispositions, structured structures predisposed to function as structuring structures, that is, as principles which generate and organise practices and representations that can be objectively adapted to their outcomes without presupposing a conscious aiming at ends or an express mastery of the operations necessary in order to attain them. Objectively 'regulated' and 'regular' without being in any way the product of obedience to rules, they can be collectively orchestrated without being the product of the organising action of a conductor. (Bourdieu 1990, 53)

The care taken here to avoid the positing of a 'conductor' having planned everything all along is important. Important too is the resistance to positing an 'express master', i.e. someone who engages constantly in 'strategic calculation' performed 'in a conscious mode' (Bourdieu 1990, 53). But whether or not planned, and whether or not consciously mediated, the world in which Bourdieu wants to picture persons as acting is a world (the 'practical world' as he calls it) 'of already realised ends – procedures to follow, paths to take – and of objects endowed with a "permanent teleological character", in Husserl's phrase, tools or institutions' (Bourdieu 1990, 53) – and it is this practical world that is said to inhere, as it were, in our 'systems of durable, transposed dispositions.' This sense of the habitus as a kind of carrier for reproducing the past (without realising one is doing so) is everywhere present in Bourdieu's account. Consider the following passages:

The habitus, a product of history, produces individual and collective practices – more history – in accordance with the schemes generated by history. It ensures the active presence of past experiences, which, deposited in each organism in the form of schemes of perception, thought and action, tend to guarantee the 'correctness' of practices and their constantly over time, more reliably than all formal rules and explicit norms. (Bourdieu 1990, 54)

The habitus – embodied history, internalised as a second nature and so forgotten as history – is the active presence of the whole past of which it is the product. (Bourdieu 1990, 56)

The body believes in what it plays at: it weeps if it mimes grief. It does not represent what it performs, it does not memorise the past, it enacts the past, bringing it back to life. What is 'learned by the body' is not something that one has, like knowledge that can be brandished, but something that one is. (Bourdieu 1990, 73)

Interestingly, for our purposes, the first of the above quotes also refers to rules and norms – but notice that it confines rules and norms to 'formal' or 'explicit' devices, thereby allowing for the transmission of the same general strategy (the same general view about how we are necessarily governed by necessarily shared determinants of what is appropriate and inappropriate), but without having to characterise those determinants as formal or explicit rules or norms (which would be implausible – on Bourdieu's own terms – for it would make us characterise the social world as the result of planning by some authority).

Again, there are passages in Bourdieu where the relationship between 'the past' and 'the present' appears more complicated than the above suggests. Bourdieu is keen, as we have noted, to make room for 'active apprehension', and thus for a certain sense of change, novelty and diversity. But what is this 'certain sense'? We get a good clue when Bourdieu comes to discuss how individual persons can have their own 'singular trajectories' (as he puts it) but nevertheless be said to necessarily share 'the same' habitus (e.g. the same 'class habitus'): 'Each individual system of dispositions is a structural variant of the others, expressing the singularity of its position within the class and the trajectory' (Bourdieu 1990, 60); and again: '"Personal" style…is never more than a deviation in relation to the style of a period or class, so that it relates back to the common style' (Bourdieu 1990, 60). Diversity, then, is only ever but a variation of something in the background that is shared: some condition of possibility, or some set of fixed enablers, which generate difference. The same strategy pervades the discussion of 'active apprehension'; Bourdieu wants to depict us as learning so as to

be able to engage in what he calls 'the intentionless invention of regulated improvisation' (Bourdieu 1990, 57): in other words, all that is 'new' and all that is 'created' is but necessarily a variation or version, or instantiation, of something that precedes it.

It is important to see that Bourdieu does not mean to refer to the mechanical reproduction of the same; he is careful to distance himself from a kind of picture where we would have to learn to do specific actions, such that we are doomed to reproducing those very same specific actions in the same circumstances. But he in fact offers a version of this, precisely because all change, novelty and diversity for him must always be a version of something prior, or something the possibility of which was inherent in certain antecedent conditions, but where that 'something prior' or those 'antecedent conditions' are described as on another plane, as a kind of general syntax that allows for infinite sentences to be produced. This is a kind of determinism because Bourdieu's sense of something in the background, at a higher level of generality (of that 'common style', or that 'class habitus') assumes what is done (in 'the present') is automatically a version / variation / instantiation of something in more general, neglecting to see that 'what is done' is up for grabs, something to be characterised as being done in some way (which is precisely what allows it to be potentially characterised as a version / variation / instantiation). Nothing is automatically a version / variation / instantiation, but Bourdieu sets things up in such a way that that is precisely what is assumed by his account; or, put differently, the idea only works if one assumes that something someone does automatically attains the status of being a version etc, i.e. of having been generated by what came before. We see that this assumption is problematic when we see that something (what that something 'is' has to be left open by the description of the case) can change what we think (what we treat as being) in that higher level, that general scheme, i.e. it can change what we think is general, enabling, prior, or 'the past'.

Of course, we can have the experience – we saw this in Klaassen et al's discussion of Wittgenstein's process of finding the 'right' word (see Section 19.A. above) – of thinking that once we settle on some characterisation of something (or just immediately treat it as being something), that it could not have been otherwise, and from there it is a small step to seeing it as a version etc that was enabled by some prior condition (existing at some general level – a general principle, say, though implicit). But what is forgotten here is that it is not a foregone conclusion that we will

see something (again 'what it is' must be left open) as either the same pattern being reproduced or a variation of a pattern that was always there. The point is that there is no determinacy here: we can see something as a novel opportunity or possibility, and do so in ways that does not always and already belong to some implicit scheme of appropriateness or intelligibility.

To make this a little clearer, consider the following example. I do not first learn a general category of 'rabbit' that I then apply (such that, in applying that category, I passively recognise instances of that category). Instead, I am exposed to particulars. My being so exposed allows me to draw, in the future, on my experience of those particulars to group them in various ways. Imagine I grew up on a farm being exposed to various animals. Later, I move to the city. On my first walk in the city I come across a pet shop, and I see something in the window. My response to that 'something' (we are purposely trying to avoid characterising it, for that would be presupposing how the agent could / would respond) could be, for example, that it reminds me sufficiently of what I now group together as 'little white furry animals that live either on farms or in city pet shops', but I could also instead see a difference (where others might see a similarity), and think of the 'something' I encounter as 'cute little white things sold by rich city dwellers to other rich city dwellers' (not seeing any similarity to animals, but perhaps seeing a similarity to what I saw happening at markets nearby to my farm). Now, of course, you could come along and tell me: 'look, these are rabbits – they are the same thing you must have seen on the farm'. We could argue about this, and I could insist (whether I would succeed in convincing you is another matter) that the relevant level of generality here is not the biological make-up of the animal, but its economic status (where it is sold, by whom to whom). What this avoids is describing the one agent as necessarily sharing the same general model as the other agent with whom the disagreement arises – indeed, without characterising it the way we have, we probably could not have made room for disagreement (and that is sometimes the point of the examples we are trying to unravel, i.e. they want to present social life as homogenous and harmonious in some perfectly ordered way). Further, it is not the case that one agent has an instantiation (a stylistic variation) of some general model, and the other another instantiation. Instead, both agents have been exposed to various particulars, which they may have, at various occasions, grouped in ad hoc ways for various purposes, but where their exposure and their groupings are not assumed to be (though they could conceivably be) the same.

When we have our disagreement, we could resolve it (though we need not, but we could) by adopting some systematisation – we agree, say, that both are rabbits, and I agree to ignore the differences I saw before. But the point of the example is that nothing determines the systematisation – what we must do, then, is avoid descriptions of cases that make it seem as if we must see things in ways that are always and already systematised for us, going to say, on the back of that assumption, that we must necessarily share those systematisations. Differently put, nothing has the permanent status of a general model that is simply applied on future instances, but we can treat, at certain times, some grouping as having a general status.

Bourdieu says that he does not want his picture to be determinist, but the problem is that his account of how we are active, and of change, novelty and diversity, is one that shows he is committed to a kind of determinism, just one that is harder to see (for it goes under the label of the allegedly more general, i.e. Bourdieu thinks he can escape determinism by saying that he does not embrace the mechanical reproduction of the same, but he fails to see that it is still deterministic to call whatever happens inevitably a version / variation / instantiation of something more general). In other words, what Bourdieu does not see (for he presupposes it) is that what has, at any moment, that status of being treated as general, is not fixed (whether metaphysically or by practical necessity), but something that can arise in novel and indeterminate ways. If we can resist this appeal to a general model, of which everything is but a variation (where all differences between persons are but stylistic idiosyncrasies), then we may be also able to resist the idea that, insofar as we live in the 'same' community we must necessarily share the same 'schemes of perception, thought and action'. Although we can, for various purposes, build structures of sameness, we are not exhaustively describable as forming groups, communities, or classes with the same dispositions.

D. The 'Creativity' of Social Action

We have seen how difficult it is to picture – and how much rides on how we picture – the sense in which we are active and creative (or how, in our sense of it, we can be different in a way that does not always and already make sense, or be classifiable as

nonsense, according to some implicitly existing scheme).[208] One of the theorists of social action attune to the issue of creativity is Hans Joas. We need to take a look at the sense in which he tries to capture – as he put it in his title – *The Creativity of Action* (1996).

It is important to see from the outset that Joas's argument in *The Creativity of Action* is not to say that philosophers and sociologists who have theorised about action (including social action) have ignored the notion of creativity. In fact, as Joas shows in the early chapters in his book (but especially chapter 2), creativity has sometimes been referred to, but when it has, it has often been presented as, at best, a kind of action – one possible way to act. In chapter 2 of the book, Joas outlines three 'metaphors' (as he calls them) under which creativity has been referred to: 1) expression; 2) production; 3) and revolution. But, as he points out, in each case what these metaphors pick out are types of actions: if we look at creativity-as-expression (as, for instance, in the work of J.H. Herder) we are likely to think of the arts, and perhaps (in romantic spirit) of individual 'geniuses'; if we look at creativity-as-production (taking our cue from Karl Marx) then we may limit the notion of creativity to productive labour; and if we look to creativity-as-revolution (here Joas looks to the work of Cornelius Castoriadis), our notion of creativity will tend to limit itself to acts charged with political fervour and ones that (perhaps from the benefit of hindsight) can be said to have instituted a political change. Although, says Joas, there is plenty to learn from these metaphors, their limits consist in the tendency to limit creativity to a certain class of actions. By contrast, what Joas wants to do is show how creativity is an integral part of all actions – or, as Joas put it, to present 'creativity as an analytical dimension of all human action' (Joas 1996, 116).

The central tradition that Joas draws on in developing his account is that of pragmatism. What is of key importance in pragmatism for Joas is the sense (on his reading) in which the pragmatists made room – within the concept of action – for 'coping with real problems encountered in the course of action' (Joas 1996, 128). The basic structure of the explanation is as follows: first, and as the default case, one pictures action as 'anchored in an unreflected belief in self-evident facts and successful habits'; second, one pictures those routines of action being 'shattered', as

---

[208] Again, this dichotomy of sense and non-sense can make us think we must always find things either immediately intelligible or immediately unintelligible, neglecting to see the middle possibility: that of the rougher ground of an inherently tentative relation, and thus one unsure as to whether and how something is intelligible or not, going on, potentially, to create a way it could make sense.

when 'the world reveals itself to have shattered our unreflected expectations; our habitual actions meet with the resistance from the world and rebound back on us'; third, there is the reconstructive phase, i.e. the 'reconstruction of the interrupted context', during which 'our perception must come to terms with new or different aspects of reality; action must be applied to different points of the world, or must restructure itself' – it being this third step that is 'a creative achievement on the part of the actor'; fourth, and finally, when and if the actor 'succeeds in reorienting the action on the basis of his changed perception and thus continuing with it, then something new enters the world: a new mode of acting, which can gradually take root and thus itself become an unreflected routine' (Joas 1996, 128-9).

As a result of the fourth step we return back to equilibrium – a kind of equilibrium that, in fact, marks the absence of creativity (given that creativity is defined as reconstruction in the face of something problematic, unfamiliar, the interrupted): 'creativity here is seen as something which is performed in situations which call for solutions, and not as an unconstrained production of something new without any constitutive background in unreflected habits' (Joas 1996, 129).

There are positive aspects to this (as Joas presents it) pragmatist account (i.e. aspects that position the mind actively), for it wants to allow for the possibility of 'new actions' or 'new solutions', but there is one key feature that retains the passivity of the mind, i.e. it pictures persons as faced with situations that themselves (so to speak) 'call for solutions'. In other words, the subject is described as faced or confronted with a scenario that in itself is problematic, that of its own accord interrupts the otherwise un-creative subject; the subject is not allowed the initiative to consider something to be problematic.

The relationship between Joas's own model and the above account (the key proponents of the above account, according to Joas, are William James, Charles Sanders Peirce, John Dewey and G.H. Mead) is unclear in Joas's book. On the one hand, Joas says he does not want to picture creativity as 'mere skilful problem-solving or even adaptive behaviour' (Joas 1996, 129), but on the other hand, he does not himself turn away from the dualism between habitual or routine action on the one hand, and moments of creativity on the other. In effect, creativity is here nestled in between large spans of habitual or routine action. The result is that creativity is not so

much integral or necessary to action as it is (at best) contingently part of it, remaining dependent on the (non-creative) habitual or routine.[209]

Joas's construal of creativity (as creative reconstruction, following what he characterises as the pragmatist approach) is in large part the effect of his wanting to escape from a conception of action he calls 'rational', which, for Joas, has the pernicious effect that it 'isolates…individual action from its context' (Joas 1996, 146). There are two ways in which Joas uses 'context: first, as the 'situation' in which an actor acts (or, better, is described as acting); and second, as the biographical context, which is designed to avoid the isolation of a single act (Joas mentions this, but in fact he places virtually all the emphasis on 'the situation'). 'Every action', he says, 'takes place in a situation'; 'the situation is constitutive of action' (Joas 1996, 160). What Joas does not want is what he takes the rational actor model to be doing, namely: conceiving of situations as 'merely a neutral field of activity for intentions which were conceived outside of that situation'; instead, he wants to see situations as the frame in which certain possibilities of actions (including evaluations of them as appropriate or not) are found appealing. Actions – possible and appropriate actions – are correlated with situations: 'every habit of action and every rule of action contains assumptions about the type of situations in which it is appropriate to proceed according to the particular habit or rule' (Joas 1996, 160).

The critical question here is how we come to find ourselves in certain type of situations. The way that Joas answers this is part and parcel of how he responds to the 'rational action' model. Thus, Joas argues that he wants to avoid the imposition of 'intentions' from the outside or, as this was sometimes put, of the imposition of 'external goals', i.e. of thinking that goals are defined in advance of action (by some isolated, independent, cogito, which directs the body; Joas 1996, 157). In other words, what Joas wants is to avoid such a conception of the relation of ends to action so as to make it necessary to split cognition and action, and make cognition antecedent to action. The way he conceives of 'cognition' is as necessarily reflective or deliberative (indeed, he also conceives of intention as consisting of 'self-reflective control which

---

[209] Others have noticed this too: for instance, Benjamin Dalton has recently argued that Joas's conceptualisation of action 'creates an unsupportable dualism between habitual action and creativity that neglects other possible sources of creative action, including habit itself' (Dalton 2004, 603). Dalton's own position is difficult to assess, for he relies so much on Bourdieu's habitus, and thus tends to picture creativity as a product of 'the interaction between (general) habit and specific, concrete situations' (Dalton 2004, 614), i.e. where something precisely general is always behind, always enabling or generating, what persons do.

we exercise over our current behaviour': Joas 1996, 158); that is why he speaks of 'pre-reflective goals', which he says are 'located in our bodies. It is the body's capabilities, habits and ways of relating to the environment which form the background to all conscious goal-setting' (Joas 1996, 158). Now it is these 'pre-reflective goals', and this 'practical-corporeal relationship to the world' (Joas 1996, 159) that also frames the world in a certain way; it is thanks to those 'pre-reflective goals' that are located in our 'bodies' (in our bodies' capabilities, habits and ways of relating to the environment) that we find ourselves in certain situations. The most important thing, then, for Joas is to avoid positing an independent, prior, calculative, deliberative, reflective cogito that imposes its intentions irrespective of the situation before it and directs the body, dictating the ends that the body has to find a means to (sometimes the mind is also depicted as dictating the means). So when Joas refers to us 'judging' or 'perceiving' the 'nature of the situation' (Joas 1996, 160), what he means is that that judging or perceiving is not done at one point prior to the situation, in some deliberative or reflective way, but rather in the course of action and thanks to the pre-reflective goals that structure our bodily capabilities and dispositions.

One might ask where the 'pre-reflective goals' come from, or how they come to be our 'pre-reflective goals', and here is where Joas draws on sociality: those goals are not 'generated by conscious intentionality' but are 'preceded' by 'a structure of common action' (Joas 1996, 184). The idea is that there is no 'pre-social, substantive Self' (a kind of isolated, rational ego, a blind will as it were), but instead that 'the individual ability to act' has a 'social basis' (Joas 1996, 184). Drawing to a large extent on Mead, Joas argues that 'Structures of social action' (one example of which is 'language', a 'collective order') 'cannot be reduced to individual actions and are thus also the precondition for the existence of interpersonal relations or social orders' (Joas 1996, 190). Joas's position is not that these pre-reflective goals that have a social basis must forever remain fixed; not at all, they are 'open to continuous revision' (Joas 1996, 161). But although they are subject to change, in any one instance of social action (or social relations) there is always an 'implicit and holistically structured background knowledge' (Joas 1996, 189-90) that makes it possible, i.e. that frames the situation for us and thus triggers the habits and rules of action that are associated with that situation.

Joas is fighting a particular fight here: the fight against what he perceives to be the rational actor model, i.e. of the positing, before action, of a cogito or will that

calculates or deliberates its way to orchestrating the body to do what it wants it to do. What he does, though, is replaces this with the socialised body. Joas wants to avoid the idea that there is a process of cognition before action. Nevertheless, he does require such a process, just not a deliberative one, but instead a 'practical-corporeal' one with determinative roots in the social. For present purposes, it is not the mode that is at issue here – not whether we judge or evaluate the situation by way of a calculative or deliberative route, or whether it is somehow automatically framed for us via a practical-corporeal route – but rather precisely that we are described as possessing a certain stock of framing devices, which are further correlated with certain rules and habits of action. It is the insistence on there necessarily being something prior that frames a situation for us as a situation, determining what is relevant, that is the trouble.

In effect, what Joas has us doing is becoming members of communities governed by rules – rules that take the form: when you find yourself in situation of type X, then apply such-and-such habits and rules of action – by virtue of our bodies absorbing those rules in the form of 'pre-reflective goals' (i.e. as bodily capacities, dispositions, etc). In other words, Joas's argument is a form of arguing for the necessity of shared determinants: the determinants are presumed to be constitutive of a certain social order or community order, to which individuals are socialised in an embodied way, such that the socialised bodies of those individuals then frame the situation for them and trigger the associated rules and habits of action. In other words: we still have here a process (not reflective, but still cognitive in the broad sense of cognitive) whereby a situation is framed, and only then action follows.

Where, you might ask, is creativity in all of this? Indeed, it is not easy to find. Presumably, Joas would argue – for he argues that we can reflect on our pre-reflective goals and revise them, leading to a new kind of habit – this would occur when we find ourselves in a situation that calls for a new solution. But just how, on Joas's own model, would we ever find ourselves in that situation? Recall that we cannot find ourselves in a situation save when it is always and already framed by our pre-reflective goals: that is the definition of being in a situation, and Joas needs that definition for otherwise he could not distinguish himself (in the way he wants to) from the a-contextual rational will standing, as it were, outside or prior to the environment. Given that, on this view, the world is always and already divided, such that we unreflectively, and in an embodied, direct, immediate manner, find it

'accessible and inaccessible, familiar and unfamiliar, controllable and uncontrollable, responsive and unresponsive' (Joas 1996, 159), one wonders how one can ever exercise, or have the impetus to exercise, creativity. It seems that, in his eagerness to avoid the controlling, reflective and autonomous will, Joas has pushed us into an extreme where the world is always familiar to us (and to all of us, in the one community, in the same way, as a result of socialisation), with the result that it is difficult to see just how it could ever be experienced by us as unfamiliar.

If one begins with the idea that there is a certain stock of situations, for one argues that it is thanks to there being situations that there is social order or community life, and one makes action dependent on situations (recall that, for Joas, the situation is constitutive of action), requiring the person to always and already be situated in one or another situation out of a limited catalogue of situations, then it becomes very difficult to see how there can be much creativity in action. Certainly, there may be something in the argument that we can be creative when faced with an unfamiliar obstacle to a task we are used to performing, requiring us, say, to invent a new means for achieving a familiar end. But this is a very narrow reading of creativity, and one that demands that the great majority of our actions be the (non-creative) pursuit of internalised ends, supported by habitual or routine means. To make room for the creativity of action, then, one must avoid the idea that 'the situation is constitutive of action', and the related idea of a passive agent finding themselves in a situation. One must, instead, return back the indeterminacy of relevance to the acting agent, i.e. one must allow the agent to actively, contingently introduce what is to be relevant (including what it might be relevant to pursue, to treat as an end).[210]

E. Enacting Social Life

We are now almost at the conclusion of our critical discussion of not only theories that speak of ways in which we act socially appropriately, yet unreflectively, but also of theories that depict us as necessarily sharing determinants of appropriateness and inappropriateness.[211] Before we conclude this Chapter we need to mention that recent

---

[210] Though it is important to see that this is still phrased in highly individual, rather than relational terms (we are doing this to follow Joas's language): the point is, in fact, probably better understood in the context of person-to-person relations where the idea is that the other, in the course my interacting with her, can create what might be relevant for us.

[211] Though sometimes only indirectly, as we saw in Velleman's work, where what we share first and foremost are determinants of intelligibility, which are – for Velleman – the foundations for appropriateness; see Section 15.B.

years have seen, in the domain of social cognition, a number of proposals, which despite drawing on innovative new ideas for imagining how we act, nevertheless still presuppose and require persons to share what are understood to be constitutive elements of social order, community life, or a culture. We will be returning to some of these innovative new theories of social cognition in Part V – they are of great importance and offer much that is fresh and new – but here we want to mention just one example of a contribution to what has become known as 'enactive social cognition', which nevertheless retains precisely the presupposition and requirement for sharing determinants of what is appropriate and what is inappropriate.

The exemplary case of how strong the hold of the governance view can be is evident in Pierre Steiner and John Stewart's paper: 'From Autonomy to Heteronomy (and Back): The Enaction of Social Life' (2009). Steiner and Stewart argue that the focus within the enactive approach on dyadic relations, i.e. face-to-face interactions, is not satisfactory. Their statement as to why offers a clear example of the idea that we must presuppose a social order that participants in that order must share in order to participate in it:

> We hold that a truly social domain is always defined by a set of structural norms; and moreover, that these social structures are not only a set of constraints, but actually constitute the possibility of enacting worlds that would just not exist without them. Interactions between two or more agents are never properly social unless they take place in the context of an environment of social structures or norms which give meaning to the interactions. (Steiner and Stewart 2009, 528)

For Steiner and Stewart, actions and interactions are not possible – will not count as actions or interactions – unless they are 'embedded in a shared normative order' (Steiner and Stewart 2009, 528). 'It is because', they say, that 'a certain behaviour is produced in accordance with norms that it constitutes an action' (Steiner and Stewart 2009, 529). To say anything else would be to '"explain away" the social' (Steiner and Stewart 2009, 528). We see also in their argument reference to the demarcation problem, such that the constitutive role of the norm also marks out what is distinctively human: 'a human being could quite well dispense with such norms… at the price of having nothing properly human to do' (Steiner and Stewart 2009, 540). Given, as we have seen above many times, that it would be implausible to argue that the norms in question are explicit (in the sense of articulated propositions), Steiner

and Stewart argue that they can be made explicit, but otherwise are and often remain explicit: 'when they are made explicit, norms may take the form of prescriptive/proscriptive/permissive statements; but most of the time they are implicit (and all the more powerful for that)' (Steiner and Stewart 2009, 536).

Once the allegedly shared normative order (composed of implicit norms, some of which have been made explicit) is in place (as a presupposition; they sometimes use the phrase 'pre-existing social order', e.g. Steiner and Stewart 2009, 530) Steiner and Stewart are then able to focus on describing how such an order is 'enacted' (hence also their subtitle, 'enacting social life') in particular interactions. Indeed, they can – and – go further, and say that the social order does not exist autonomously from, free-floating above, interactions but rather 'inheres' in those interactions (Steiner and Stewart 2009, 538). But here talk of 'inherence' side by side talk of 'pre-existence' is like trying to have one's cake and eat it: Steiner and Stewart want to keep the idea of a pre-existing shared normative order, but seeing that they are also committed to the idea of enactive social cognition, which places so much emphasis on dyadic interactions, they try and infuse those interactions with that order, making it seem as if the driving force are the interactions. Steiner and Stewart, then, depict interactions as the carriers or transmitters of an order – enacting and, in the course of enacting, reinforcing the order – that they presuppose. The problem for them is that having begun with the presupposition of a shared normative order, there simply is no other way for them to describe interaction (talk of 'inherence' is an attempt to deflect attention from the way in which interaction is imagined as simply enacting and re-enacting that presupposed order).

Steiner and Stewart speak repeatedly in their paper of 'sociality' (in the sense of shared normative order) being 're-actualised in some situation', meaning that 'the individuals concerned [in the situation] cannot but produce their own behaviours on the basis of their implicit knowledge of the historically formed norms or structures that define what ought and what ought not to be done in that particular situation' (Steiner and Stewart 2009, 537). Notice the perfect fit here between the way that Steiner and Stewart conceive of sociality – as shared normative order (and thus constituted by certain norms) – and the knowledge imputed to participants; the moment one questions this, by for instance suggesting that any particular example of interaction is not necessarily mediated by knowledge of the normative order (say because it seems implausible to think that some domain of interaction is so tightly

regulated as to be specifiable as an order), Steiner and Stewart refer to there being implicit norms known implicitly and processed subpersonally.

There are two reasons why Steiner and Stewart make the above move (i.e. presuppose, and then describe persons learning and interacting on the basis of their implicit knowledge of, a shared normative order). The first is that they think it would be implausible to picture persons as instituting or planning sociality from the beginning, from square one as it were: that is why Steiner and Stewart keep referring to social structures that 'are external to the individual' and which 'existed before he was born, and…will survive after he is dead' (Steiner and Stewart 2009, 537). The second and related reason is that picturing the individual in the position of instituting or planning sociality from scratch would also result in an implausible picture of cognition, i.e. of a completely a-contextual, a-historical picture of the individual (and of individuals interacting). But what Steiner and Stewart do not see is that the alternative to a shared normative order is not that: it is not bare individuals starting from scratch. One can – and indeed must – depict persons as learning: their abilities to act and interact do not come from nowhere. The key is how we picture them learning: do we presuppose some specific set of things that pre-exists them that they learn and then reproduce or re-enact, such that members of the same group learn those same things thereby allegedly enabling them to interact together, or do we allow for something more flexible and also heterogeneous – do we allow, in other words, for both 'what' is being learnt and the mode of learning to be indeterminate (thereby also rescuing diversity, but without giving up on the possibility of sharing – just not making sharing necessary)? Throughout this thesis we have been trying to make room for precisely this alternative perspective, and we will be systematising this and giving more examples in Parts V and VI.

It is doubtless unnecessary to add that we have not had occasion here to cover all the various ways in which persons are depicted as being able to act socially appropriately, yet unreflectively. For instance, we have said nothing about the way in which emotions are quite often depicted as playing the role of 'enforces' of social norms, as in Jon Elster's work, where the very anticipation of the emotions of anger, guilt or shame is said to enforce and indeed reinforce social norms (Elster 1994, 31); or where they are said to be the carriers of the 'compellingness' of that which is necessarily shared, as in the work of social psychologists Cor Baerveldt and Paul

Voestermans (2005).[212] From the second-person perspective, social emotions need not be conceived in ways that make them function as a kind of internalised police for social norms. This is, once more, precisely because we do not need to begin our study of social life (including its normative dimension) with the presupposition of necessarily shared norms or other determinants of what is socially appropriate and inappropriate. Throughout this thesis so far we have been showing why this strategy does not work, and how it makes us think not only of social life in misleading ways, but also paints us (our minds) as overly passive and rigid. We now need to systematise what falls through the cracks of the governance view, i.e. we need to give a positive account of the second-person perspective on social life, including on its normative dimension.

---

[212] In more recent work, Baerveldt has combined with Theo Verheggen (2007) to argue against the necessity of sharing, but he replaces this with 'domains of consensual coordinations of actions' (see, e.g. Baerveldt and Verheggen 2007, 17) – in other words, the key issue for him is the mode in which we come to act in the same way, where that mode is not necessarily understandable as 'states of propositional agreement'; what we have been arguing here is that it is not the mode that is the issue, but precisely requiring sameness (or in his language, 'consensus').

Part V. The Second-Person Perspective

We have now progressed through four Parts in each of which we have looked in detail at an element of what we have here called 'the governance view'. According to this view, the normative dimension of social life consists in the way in which determinants of appropriateness and inappropriateness (which were, in general, discussed in Part I) are: 1) necessary (the first element, discussed in Part II); 2) necessarily govern the mind (the second element, discussed in Part III); and 3) are necessarily shared by us (the third element, discussed in Part IV). It has been argued with respect to each of these elements that we have good reasons to think that the governance view over-states, over-emphasises, or over-generalises the importance of normative resources – as we have seen above, from the perspective in which they are not over-stated, etc, they are normative resources that are contingently useful, whereas the governance view thinks of them as determinants of appropriateness and inappropriateness that are necessary.

Where the governance view wanted necessity, we have, with a view to making room for the second-person perspective, introduced contingency: we *may* draw on normative resources to guide ourselves or to evaluate ourselves and others, where such 'drawing on' is an active treating of those resources as relevant (including, possibly, treating what they require as unchangeable, and thereby experiencing those normative resources as necessary) and not something applicable and applied either by metaphysical necessity (e.g. as is often claimed in the context of logic) or by practical necessity (e.g. as is often claimed in the context of grammar or chess).

At each turn, then, we have tried to show what it is that the governance view misses, i.e. misses as a result of how it construes the normative, including the normative dimension of social life. In particular, we have seen how the governance view tends to position minds-in-action (individual minds) and minds-in-interaction (social minds) in a passive position, and thus unduly narrows our understanding of how minds learn and work (again, we are using 'minds' broadly here, as in the term 'philosophy of mind', without meaning to indicate or endorse any dualisms). The reader may have felt this to be an overly negative exercise, but it was a necessary one: without carefully proceeding through the elements of the view, we might have left it open for adherents of that view to argue that we have neglected to see how

determinants really are necessary, how they necessarily govern us, and how they are necessarily shared.

Having so proceeded through the governance view, and, at each point in the critique, also tried to offer glimpses (including examples) of what was being missed, it is time to systematise what has throughout this thesis been referred as the second-person perspective. Of course, given space limitations, no attempt at a comprehensive and full coverage can be made. The aim of these next two, final parts of the thesis is to systematise the glimpses we have offered and to sketch an outline that will be sufficient for opening up prospects for further work.

The first of these, the current fifth Part, outlines 'the second-person perspective', or, perhaps more accurately, offers an account of how social life is understood from the second-person perspective. The second-person perspective is, primarily, a perspective on the social in general: it brings into view a way that interaction can be experienced by persons, and it sets the scene for the description of abilities (undertaken in the sixth and final Part) that persons need to participate in the social (precisely conceived from the second-person perspective). Although the second-person perspective is primarily a perspective on the social in general, it also has implications for our understanding of the normative dimension of social life. In part, the implications are negative: if we are to take what the second-person perspective brings into view seriously, then we cannot adopt the governance view of the normative dimension of social life. This is because, as we have seen, the governance view squeezes out the possibility for the second-person perspective. In part, however, the implications are positive: the second-person perspective brings into view not only experiences of social life in general, but also their normative aspect, i.e. it enables us to see and describe, in a certain way, how persons experience each other normatively in the course of interacting with each other, and (less subjectively) how being able to relate normatively with each other in those ways (having those normative experiences of each other) is important for persons being able to participate successfully in social life. Put another way: the second-person perspective brings into view certain normative experiences we have of each other, in the course of interacting with each other, that the governance view of the normative dimension of social life does not, and cannot, recognise, or at the very least finds it difficult to acknowledge as important.

More concretely, then, what is vital about social life from the second-person perspective is that the other person need not be imagined or presupposed to be the same; indeed, acknowledging and maintaining difference is critical for interaction understood from a second-person perspective (though difference here has to be understood in a specific way). This has interesting effects on how interaction is perceived and described. Of particular importance is that it allows us to think of interaction as falling in and out of balance, as partly hesitant and partly confident, as partly fluent and partly aberrant; in short, it allows us to avoid positing a particular underlying pattern or rule – something that conditions the possibility of interaction. What is critical, then, is that the second-person perspective allows us to think of interaction as incomplete, imperfect, asymmetrical, awkward, always somewhat askew; this may not seem very neat and not nearly as theoretically satisfying as the idea of an interactive or social order (because it does not give us a neatly defined object of explanation), but that is precisely its advantage and strength. In other words, methodologically, making room for and recognising the importance of the second-person perspective allows us to avoid constructing a target or object of explanation that requires us to explain (and typically reify, or at least over-emphasise) a certain set of regularities or a certain set of determinants of appropriateness or inappropriateness. The second-person perspective allows us to see that social life – including in its normative dimension – is not a patterned order, or an order of regularities with attached normative attitudes, or a set of rules that becomes internalised and regularised (though it also shows that sometimes, in certain contexts and for certain purposes, we can treat it as if it was). It shows us, instead, that social life, again including in its normative dimension, is a mix of the regular and the irregular, the partly predictable and partly unpredictable, the fluent and the awkward, and it further shows us that seeing this about social life is vital for seeing and describing a fuller array of the kinds of abilities persons need to participate in social life successfully (including having the fuller array of normative experiences of other persons – precisely those that are not convention-, norm- or rule-based – that persons need to have in order to, once again, interact with them more successfully than they might otherwise).

Later, in Part VI, the thesis will turn to describe the abilities in question. In short, Part VI will argue that there is a distinctive kind of cognition – second-person cognition – which persons need to participate in social life, at least as long as we

recognise what the second-person perspective brings into view about social life. This cognition is, as we shall see, comprised of a partnership between what is called 'appreciation' and 'attentive silence', both technical terms that will be fully elaborated on in the two Chapters of Part VI. We need, however, to proceed step-by-step, turning, in the first instance, to articulating what it is that the second-person perspective brings into view about social life.

There are four Chapters in Part IV. The first two Chapters draw on the work of two theorists: Vasudevi Reddy and Hanneke De Jaegher. We shall see how Reddy and De Jaegher have constructed (or at the very least help us to construct) a picture of interaction where sameness is not presupposed or required by necessity, and where difference (itself understood in a certain way) is kept track of. Reddy's work, considered in the first Chapter of this Part (Chapter 20), offers a perfect introduction to interaction understood from the second-person perspective. Importantly, Reddy places a good deal of emphasis on play, though here, as we shall see, we need to be careful and not allow play to be understood as but the internalisation of rules that that play is argued to be structured by (that is, in effect, to adopt a governance view approach). De Jaegher's view, meanwhile, discussed in the second Chapter of this Part (Chapter 21), will help us get clearer on the kind of asymmetry, incompleteness and imperfection we need to see in what she calls 'the social interaction rhythm'. Our aim, it should be said, is not to defend Reddy and De Jaegher's views, but to present selected aspects of their views and see how they can help us get clearer about what kinds of insights the second-person perspective can offer.

Our next step in articulating the second-person perspective, undertaken in Chapter 22, is to see how we can conceive of interaction breaking down. We will draw on a very different understanding of 'breaking down' than is promoted by the governance view (where breaking down is defined in terms of the breach or violation of some determinant of what is appropriate). In this respect, we will look principally at studies of mother-infant relations. What is interesting about these studies is that it is precisely the need for fixing how the interaction ought to proceed that places the (quality of the) interaction at risk. It should be stressed, at the outset, that the thesis does not pretend to, or hope to, making a contribution to the extensive literature on mother-infant relations. Neither does the thesis attempt to present the little of what is discussed in this Chapter as authoritative, within the psychological literature, about those relations. Instead, the thesis uses that context, and some of that literature, to

make a conceptual point, i.e. a point about how to conceive of interaction going more or less well from the second-person perspective. To repeat, from the second-person perspective interaction goes less well when it becomes so regularised, or so structured by rules that are on the road to becoming regularised, that there is virtually no room for irregularity, unpredictability, uncertainty, ambivalence, and even awkwardness.

Finally, in the last Chapter of this Part, Chapter 23, we will return, and consider in some more detail, the sense of otherness or difference that has been noted, above, as being so vital for the second-person perspective. We will see that the conception we need is not one that pictures the other as different in some specific respect, but as precisely different in a way that acknowledges the other as their own autonomous source of meaning and initiative, and thus a difference that always partly eludes our grasp – a difference that always leaves a remainder that is at once both unfinished, and partly because it is unfinished, offers an invitation or opportunity for us to respond.

There is, no doubt, more to say about the second-person perspective than is said in the four Chapters of this Part. The thesis acknowledges that there is a long tradition of thought that emphasises difference, rather than sameness, in relations between persons. Some of the relevant literature, not discussed in this thesis, has been mentioned in the Introduction to the thesis. As is noted there, in future work, that literature can fruitfully be harnessed to further develop the second-person perspective. For the moment, we will need to satisfy ourselves with making room for it and making it initially plausible and, perhaps most importantly, at least sketching out why it is important to develop it further.

Chapter 20. Infants at Play

One of the domains in which the second-person perspective has recently gained momentum is developmental psychology. Perhaps the leading contribution in this respect has been Vasudevi Reddy's. Her work is the focus of this Chapter. Having presented some of the key aspects of her work, we will then be in a position to contrast her approach with other accounts of children at play (and the importance of play for development, including moral development).

A. A Second-Personal Developmental Psychology

In her *How Infants Know Minds* (2008), Vasudevi Reddy develops an alternative to the dominance of first- and third-person 'solutions' to the old problem of understanding other minds.[213] As noted above, her specific interest is developmental psychology, but her remarks and examples are generalisable. The alternative Reddy develops raises questions about what we might take 'the problem of other minds' to be about (what is at stake in it), but let us first focus on her criticism of the dominant solutions.

According to Reddy, the first-person account 'argues that other minds are known through reference to the self… you look across the gap [between you and others], see another body, and recognise in that body a similarity (or something of) yourself' (Reddy 2008, 18). Reddy considers a number of versions of the first-person account. For instance, she recalls how it was once referred to as 'the argument from analogy', identifying John Stuart Mill as an early exponent:

> Mill describes the process by which, for example, we understand that another person is angry: we see that person making vigorous movements with the fists and stamping movements of the foot, and we recall previous occasions when we had made the same movements and had experienced a feeling of anger. We infer, through analogy with our own experience, that other bodies must also have the same sort of mental states we have. (Reddy 2008, 19)

Reddy criticises the argument from analogy in a number of ways: first, she argues that if our relation to others was indeed based on analogy, and assuming analogy is

---

[213] It is important to note that we are not here adopting Reddy's use of, or understanding of, 'first-person', 'second-person' or 'third-person' accounts / views / perspectives. As noted in the Introduction, this thesis means by the second-person perspective a way of seeing what is importantly part of social life; not a perspective a person can take in the course of interacting with another. For the purposes of discussing Reddy, however, her terminology will be adopted.

understood as working 'by extension from at least a few proven cases', then it looks like when we do employ analogy (on the basis of our experience of ourselves) then 'we are making a huge overgeneralisation from one case'; and second, she asks: 'If our experience really was so solipsistic, why would we even be tempted to make the generalisation that other minds existed?' (Reddy 2008, 19).

Reddy also mentions the contemporary version of the first-person approach, namely the simulationist account. She recognises that simulationist accounts have become very sophisticated in the last two decades, and sometimes involve elements of the third-person account. The nub of her dissatisfaction with these theories is that they still require an inference or perception of 'similarity', of seeing 'the self in the other', which she argues is unsatisfactory because 'A perception of similarity is not enough to provide the motivation for engagement and communication or to explain responsive emotions that fuel engagement' (Reddy 2008, 21). 'A first-person approach', she says, 'cannot convincingly explain interaction or dialogic engagement. Dialogue needs more than a recognition of similarity – it needs a recognition of difference and the ability to respond as well' (Reddy 2008, 25).

What we can see Reddy opening up in her critique of the first-person approach is the crucial role played by difference – and thus neither an other who can only be related with by being understood on the model of one's self, nor an other who, in the absence of such an understanding, is experienced as negatively meaningless. In other words, Reddy points to a need to see the difference between persons as a positive thing – as precisely that which enables interaction and allows it to continue and flourish, not that which is an obstacle to it, and thus something that must be overcome.

On the third-person account, the mental lives of others are presented as inaccessible to us. In order to solve the problem (of other minds), we develop hypotheses or theories (folk psychologies), and apply them to the behaviour of others we observe, inferring the presence of such-and-such a mental state or series of mental states. Examples of such hypotheses or theories include: 'that weeping indicates depression', or 'that avoidance indicates dislike' (Reddy 2008, 22), and we apply such theories to make others' behaviour intelligible to ourselves, e.g. when we see them weeping, we infer that they must be depressed. Known today as the theory of mind theory (or theory-theory) for short, Reddy argues that this popular approach assumes that 'we need a rational deductive route to minds: that some abstraction (whether a

concept or a hypothesis or a theory) from otherwise meaningless sense data is necessary to derive the idea of the psychological' (Reddy 2008, 22). Before a scheme is imposed, the other's behaviour is said to appear 'as a jumble: chaotic, disordered', meaning that 'the theory (or other reflective realisation) is seen as a prerequisite for engagement with other minds' (Reddy 2008, 22).

This way of characterising the third-person account itself raises problems (not addressed by Reddy): for instance, if the idea is that we first notice someone weeping, and then, on the basis of our theory infer that when someone is weeping they are depressed, then the question is how we can know that someone is weeping? Surely, seeing that someone is weeping is not something chaotic or disordered, but if it is not chaotic or disordered, how is it possible? Does not the third-person account (at least on this brief reconstruction) assume that we have a certain kind of access to others (in the above case, that we can see them weeping)? The third-person advocate could say here that 'weeping' is just a term that is triggered when we apply the entire hypothesis; before that, what we see are, say, tears running down a cheek. But here we could repeat the question: is seeing 'tears running down a cheek' seeing a jumble, something chaotic and disordered, and if not, then how come we are able to see that? The problem here for the third-person advocate is that one cannot *both* postulate that in the absence of the application of a hypothesis the other's behaviour is chaotic / disordered *and* maintain the idea that a hypothesis is applied. One cannot make these two postulations because the agent must first think / feel – however one explains this – that the hypothesis is relevant and applicable, and thus the agent must have some kind of access to the basis upon one which the agent decides the hypothesis is relevant and applies it, the point being that on the third-person view we have no explanation of how one can have access to this basis, for anything that is not accessible by hypothesis is said to be chaotic / disordered.

Reddy fastens onto another kind of criticism of the third-person account. She argues against it on the basis of its commitment to a detached relation to others, which, she argues, has the effect of failing to capture 'subjective experiences', such as 'the feeling of redness', as well as neglecting to see the possibility that persons can have 'access to others' experiences through often different responses to them' (in other words, the theory-theory cannot explain 'the experience of emotional engagements with others', which is vital because, for Reddy, it is via these emotional engagements with others that we can have access to others' experiences; Reddy 2008,

23-4). Reddy argues further that the third-person's commitment to detachment – prioritising, as it does, the perspective of the observer or spectator – may also be highly 'ethnocentric', attributing generalisations from one's culture (one's own culture's folk psychology) onto others (Reddy 2008, 24-5; see also 232-236).

Before we consider how Reddy distinguishes her account from both first-personal and third-personal accounts, there is an observation we need to make. The problem to which first and third personal accounts are purported solutions is one that concerns access to – or knowledge of – other minds. In trying to move beyond the first-personal and third-personal accounts, one may wish to question what these accounts understand the problem to be, and this is what Reddy does. But there are two different ways to question the understanding of the 'other minds problem': first, one may say that (at least ordinarily) there is no gap between minds, i.e. that persons ordinarily have access to – or knowledge of – each other's minds, at least when certain things go well, e.g. there is emotional engagement (or some other condition); or, second, one may say that what the problem demands (what it asks for) does not, in fact, need to be demanded, i.e. that we do not need 'access to' or 'knowledge of' other minds.

Of course, a lot hangs, in terms of the plausibility of this second way of questioning the problem, on just how the terms 'access to' and 'knowledge of' are cashed out. We will be elaborating on this as we go along, but let us note for the moment: 1) that it may be that the second-person perspective can help us see that when we relate with another person we need not, and to some extent even cannot, claim, assert or immediately notice (in a stop-and-start fashion) that someone is doing such-and-such a thing (e.g. weeping) or is in such-and-such a state (e.g. depressed), but that we may instead (need to) relate with others in more inherently tentative and flexible terms, *attending to them silently* and thus without committing to seeing how they 'really' are; 2) that even when we do *appreciate* someone appearing, say, depressed to us, this is not the application of a general model to a particular case (and thus a potential 'ground' for a claim or assertion about how we 'know' or 'have access to' how someone is feeling or what they are thinking / doing), but the exercise of a more inherently tentative and flexible ability, and one also that remains open to one's feeling about the other being changed in the course of interacting with the other.

The point is that on the first way of questioning the problem of other minds, we miss out on step 1, and even though we may then say that we can see someone is

depressed without employing a hypothesis (based on the rule mentioned above) the door is open for the advocate of the rule-based approach to say that the process for applying the rule is sub-personal (so we end up reverting to a kind of rigid description). This is why we need to adopt the second way of questioning what the 'other minds' problem demands. Of course, adopting the second way requires elaborating further on what exactly is meant by 'attending silently' and 'appreciating' and these are precisely the tasks undertaken in the next Part.

There is some ambivalence in Reddy's way of questioning the problem of other minds (the 'gap' problem, as she sometimes calls it). Sometimes, her account reads like an endorsement of what was described above as the first of the two possible ways of questioning the other minds problem, and sometimes what she says is more open to something like the second way. Consider, for instance, how she presents the 'three core features' of her approach:

> First, it rejects the 'gap', the dualist assumption criticised in the last chapter, which portrays other minds as opaque to perception, only speculatively accessible through various mechanisms of inference, modelling, or theorising about behaviour. It sees minds as transparent within (and within the limits of) active, emotionally engaged perception. Second, it pluralises the other, rejecting the assumption of singularity in the way we sometimes talk of other minds – of '*the* other'. Since we can have many different kinds of engagement with others, what we can perceive of them must vary with the degree and type of engagement… Engagement in the second person allows us to experience others within our emotional responses to them as particular others – an experiencing which is more than simply a recognition of their similarity to ourselves. Third, it sees this active emotional engagement between people as constituting – or creating – the minds that each comes to have and develop, not merely providing information about each to the other. (Reddy 2008, 26-7)

Such are the core features of Reddy's second-person approach, but in fact everything here rests on the details. For instance, it matters a great deal how the difference of others is conceived, e.g. is it conceived (as might be suggested by the term 'plural') as being possibly one out of many (but a limited many) kinds (such that a person chooses one out of fixed alternatives to apply), or is it a difference that does not belong to a stock of kinds one might already have? Similarly, what is meant by 'particular' others (are these particular in some specifiable sense, or particular precisely because they cannot be specified without loss of particularity)? Further, when persons are described as actively engaging with each other emotionally, what

does this mean – does it require persons to feel 'what' others are feeling, or does it allow for something more open, something that reads less like a claim or an assertion that can be backed up by a justification and more like an unfinished and imperfect relation with another? How are cases of persons actively engaging with each other emotionally described? We must, then, delve into the details of Reddy's account a bit more (we will still need to be very selective, but it is important to go into some more details). Doing so will help us carve out the distinctiveness of the second-person perspective.

When Reddy describes her approach in general, she does so in large part by stressing what she calls 'active emotional engagement' with others, which requires 'active emotional responsiveness' (Reddy 2008, 30-1). 'The nature of our engagement with others', Reddy asserts, 'is absolutely fundamental to our experience of them' (Reddy 2008, 30). Sometimes, when speaking of 'our *experience* of others', Reddy speaks of 'perception', and so she argues that in the course of engagement with another the 'gap between one mind and another disappears and is replaced by a perceptual link consisting of active emotional responsiveness' (Reddy 2008, 31). Reddy understands 'all perception' as 'embedded in living and doing', and not a matter of perceiving from a distance, in a disinterested, spectatorial, theoretical, observational manner, but as integrated with action and involvement in 'everyday practical life' (Reddy 2008, 29).

When she speaks in these terms, Reddy appears to stress the sense in which she thinks we can have direct access to or immediate knowledge of another, 'without…a more distant observation through the filter of reflective evaluations, anticipations or theorising' (Reddy 2008, 28). This is problematic from the perspective of this thesis, for we want to argue that it is not the case that making room for active emotional engagement with others means that such an engagement is precisely the way in which persons might have access to or immediate knowledge of another. This is problematic for our purposes because, as we have been suggesting (and as we shall see in more detail in the next Chapter when looking at De Jaegher's work), it may be that interacting with others (including actively and emotionally) requires some imbalance and uncertainty, such that it is over-idealising relations with others to make it out as if a necessary element (and benefit) of any active emotional engagement is direct access or immediate knowledge of another.

Now what is interesting about Reddy's work (and this returns us to the ambivalence mentioned above) is that in other passages when describing the second-person Reddy does acknowledge the importance of something quite close to that imbalance and uncertainty. Consider the following: 'But engagement has its limits: disengagement is not only inevitable, it provides a valuable dimension to knowledge that is born with engagement' (Reddy 2008, 34). Reddy adds, for instance, that 'Many different human skills require…[a] mix of engagement and disengagement… Humour, for instance…depends on the rapid fluctuation of disengagement and engagement…which is practiced even by babies before they are 1 year old' (Reddy 2008, 35).

'Disengagement' here has several different readings. One is a rather narrow reading – the one we want to avoid here – that speaks of disengagement as a reflective exercise of the third-person, a stepping outside of the interaction, and thus conceived of as something negative, or if valuable then valuable only as an impetus or (reflective) break from the meaningfulness of engagement (and thus something the value of which is derived from and supports engagement). That Reddy sometimes thinks of disengagement in this way is obvious not only from the above quote (speaking of 'knowledge' being 'born within engagement', and disengagement only providing 'a dimension' to that knowledge), but also when she makes observations about disengagement like the following: 'something like the sadness we might experience when realising that the complete harmony of a new love has been broken by a disturbing and unshared (or unshareable) thought' (Reddy 2008, 35; our point is that to make room for a second-person account one must drop this idealisation of 'complete harmony').

There is, however, in Reddy's work also a second reading of disengagement, and one that does not conceive of it as outside the interaction, but as part of it, as when – in another of the quotes above – she speaks of humour involving 'the rapid fluctuation of disengagement and engagement.' It is with respect to this broader reading of disengagement, which is understood to be part of, and not a break from, interaction, that we may make room for that sense of imbalance and uncertainty, and thus also avoid idealising interaction understood from the second-person perspective, or making it out as if by virtue of engaging with others we can guarantee some safe knowledge of, or un-doubtable access to, another. In other words, what we want to look for in Reddy's work is awareness of how relating to another – an 'other', whose

difference to us is always maintained, though not a specific difference but an open difference (a difference we do not quite / not yet understand) – is not (ideally) a matter of 'complete harmony' where we have 'direct access' and 'immediate knowledge', but rather is an unfinished and uncertain exploration or orientation, something that neither of us control but to which we both contribute. For this, we might need in particular to look for cases where 'disengagement', despite what Reddy sometimes says of it in general, is not understood exclusively as a matter of reflection – of standing back and re-assessing or restating – but as part of the way in which we are involved in an active, bodily and emotional way.

The first context in which we see Reddy making room for something close to that imbalance and uncertainty as part of interaction (from the second-person perspective) is in her discussion of imitation (in chapter 4 of her book). One can see how over-emphasis on engagement (as distinct from disengagement) might mean that imitation would offer a perfect beginning: the first spark of contact, and one, furthermore, that involves two persons doing the same thing (though one being passive and the other active, i.e. one imitating the other). The fact that Reddy does not read imitation in this way – preferring to think of it as a 'conversation' (Reddy 2008, 43) – suggests that it is better to read her as not thinking of engagement *versus* disengagement, but of engagement *and* disengagement in the course of an interaction. Even, then, at this basic imitational stage, the second-person perspective requires that difference between persons interacting with each other is maintained. Another way to put this is as follows: imitation might, at first, blush lend itself to support a view that reduces interaction to some sense of sameness, e.g. either by saying that I can only relate to another by thinking of them as like me, or by saying that I can only relate to another by thinking of us as the same (because we share some scheme for making each other intelligible). That Reddy does not understand imitation in this way shows us what is unique about the second-person perspective. Reddy argues, in relation to a series of studies of neonatal imitation, that 'simply explaining the infant's recognition of similarity between self and other is not enough' (Reddy 2008, 58). The following is a telling passage:

> When interacting with people newborn infants don't just imitate, they respond. They respond with interest or disinterest, with attention or avoidance, and at least within weeks, with reciprocal rather than imitative emotional actions. It has been argued that even imitative interactions are not merely imitative and

that the actions are never identical to those observed. They often involve (even in neonates…) varying degrees of approximation and varying degrees of hesitation and change: *responses to*, rather than *mimicking of*, another person. (Reddy 2008, 59; original emphasis)

There may be some confusion of terminology here: Reddy sometimes refers to 'imitation' as 'mere imitation', or of the infant in the interaction doing more than imitating; but sometimes she presents her views as being about how we understand imitation. In the second sense, her point above is that we ought not to understand studies of neonatal imitation – we ought not to interpret the data of these studies – as a matter of a baby simply repeating, or trying to repeat, exactly the same action (when we say this we are in fact assuming what the same action is, e.g. 'tongue protrusion'); instead, we ought to see the infant as attempting to respond to, interact with, another as an 'other' (in this sense, the idea is that to think of imitation as mere imitation is to reduce imitation to mimicking, and that, in turn, is to misread – fail to make room for a second-person perspective on – what infants do when they interact with adults).

Speaking, then, of imitation (and not mimicking), we might ask: why does the infant imitate? And here Reddy answers that 'The neonate, in imitating, is responding to this communicative attempt by the other by the completion of the other's as yet incomplete interpersonal action' (Reddy 2008, 61). This view is backed up by data of imitation with chimpanzees: when the experimenters used a 'face-to-face interactive setup (and not just the perception of the modelled act', and thus a 'more conversational and less rigid experimental procedure in eliciting imitations'), there was a higher rate of imitation (again, not mimicking, but responding) amongst chimpanzees (see Reddy 2008, 61). Another way of making the same point is to ask, as Reddy asks: '"Would neonates imitate tongue protrusions outside of engagement?" and even 'Would parents protrude their tongues to neonates if they weren't trying to engage them (however bizarrely)?"' (Reddy 2008, 63). Presumably, one could test this with showing pictures of persons with protruding tongues to infants, and contrasting this with 'live adults' protruding tongues.

What we see emerging here is the imitation scene as an early mode of making contact – we could see this, reductively, as a matter of simply mimicking (we could pick out an 'it' that the baby and the adult are both repeating), and thus we could subject this interaction to a description of the same thing being re-enacted (laying down the foundations for more complex re-enactments later), or we can see it (from

the second-person perspective) as the first, somewhat hesitant (though also strangely pleasing: see Reddy 208, 63) experiences of relating (responding) to another person in an active way, a kind of testing of whether the other is an other that also responds to one's self as an other. In a sense, one could say that imitation is a procedure for creating one's self as an other, precisely by seeing how / testing whether an other treats one as an other (again, the crucial thing here is difference, not sameness).

Even clearer than imitation are Reddy's analyses of dialogue or conversation. When we think about dialogue, says Reddy – whether it be anything from 'a jazz player's seamless conversations with his fellow musicians' to 'a baby and mother face to face, each enjoying an intense wordless exchange' – we see (at least from the second-person perspective) that 'there is at least some mutual recognition of the consciousness of the other, some kind of openness in each partner to the acts of the other and a taking of turns in responding to them' (Reddy 2008, 67). A dramatic case of the need for ongoing mutual recognition of otherness can be seen in what are known in the developmental literature as 'still-face' experiments: what one sees in these is how infants who were a moment ago interacting actively with an adult, now, when faced with the same adult who (only with great concentration, and a lot of difficulty) does not respond but maintains a still face, look bewildered, fidget and soon begin to cry (see Reddy 2008, 73-4). Lynne Murray famously performed a version of this experiment in 1975:

> She got 2-month-olds to interact with their mothers through television monitors. That is, the baby was look into a monitor in which his or her mother's face could be seen, live, and the mother was, similarly, looking into a monitor that framed her infant's face. The two partners were then encouraged to interact as usual; despite the strangeness of the technical mediation, it was clear (judging by the vocalisations and facial expressions of both) that perfectly happy interactions could occur between mothers and infants. Importantly for the procedure, the behaviour that each partner displayed during these live interactions could be surreptitiously videotaped. So, once happy interaction had been established (and taped) for a while, the videotape would be surreptitiously rewound and, unbeknownst to both infant and mother, replayed to them. Instead of seeing the mother live, the infant would now view the other as she had been a few minutes earlier. The infants showed puzzlement and confusion, looking back at the video image of the mother now and then as thought to heck. They showed fewer smiles, looked away more, exhibited more closed-mouth expressions, and intermittently attempted to regain interaction. In short, they could detect the changes in the mother's behaviour. Even when the mother was smiling and chatting – the very same smiles and conversational offerings form the mother that had minutes

previously produced happy responses in the infant – the infant looked sober and wary. (Reddy 2008, 75; see Murray and Trevarthen 1985, based on Murray 1980)

We have included the description of the experiment in full here, as it is important for the second-person perspective, and it is also one to which we will be returning later when looking at pathologies of mother-infant relations (in Chapter 22). The point is that the interaction is a pleasant and interesting one (or pleasant because interesting) as a result of the unpredictable difference between infant and mother, i.e. the quality of the interaction consists not in re-enactment of the same (as might be the case if the infant just played out the same scene with the mother in the videotape), but rather in the maintenance of difference (at least in part by way of experiencing one's own otherness – and thus one's self – in or with the other). In other words, 'engagement' here cannot be a matter of something predictably safe, but rather something precisely hesitant and full of change – full, if you like, of difference and uncertainty.

Notice, for instance, that the infant takes no pleasure or comfort in recognising the mother's expressions: the point, once again, is not about recognition; the child is not looking for a particular kind of reaction from the mother, such that when he receives it, he expresses acknowledgement of it as appropriate in some specific way; instead, appropriateness here consists precisely in the contingency of response, in things being able to go in a direction not controlled by one party or enabled and thus required by some external structure, with both persons participating in and contributing to how they could go differently. Reddy speaks of this by referring to the 'openness' of dialogue: 'Dialogue', she says, 'cannot be scripted and predetermined. It must possess within it the possibility of going down a road which none of those engaged in it could have known about' (Reddy 2008, 82). Drawing partly on Daniel Stern's work, Reddy points to the need for 'lived stories', i.e. a process in which 'things are open, the expected and the novel can happen, and whether rewarding or conflicting, they demand resolution and explanation' (Reddy 2008, 83; see Stern 2004).

The openness of dialogue has a counterpart in terms of the abilities of persons to participate in it: what it requires, says Reddy, is a kind of playfulness. Reddy refers to this in a number of different ways, e.g.: 'honesty in listening and responding to the other, and courage to be open to unexpected outcomes' (Reddy 2008, 83); and 'emotional sensitivity to the unpredictable' (Reddy 2008, 84). There is also

'playfulness' here in a more explicit way, as when infants engage in teasing others, provoking them or surprising them (Reddy 2008, 84) – presumably precisely as a very obvious way of testing their responsiveness to contingency and openness to things being otherwise. All this, as Reddy notes, contrasts starkly with those accounts of communication that make it necessary for relations between persons to be mediated by their mutual recognition and enactment of some sense shared script or code (see Reddy 2008, 86-7). Of course, this does not mean that scripts or codes cannot be used, just that they cannot be conditions of possibility for an interaction; indeed, interaction understood this way means that even when they are drawn on, this must be done in a tentative way (a way that respects the capacity of the other to resist the code, or change it, or reveal its limitations or irrelevance).

Playfulness features very strongly in Reddy's book. We cannot – and need not – give a full account of Reddy's treatment here, but we can note a few important instances. For example, in a chapter on infants' perception of intention (chapter 8), Reddy speaks of how intentions need to be 'engagable' by the infant and how the infant needs to engage with them. At one point she offers the following statement and set of examples:

> There is an openness – an incompleteness – in intentional action which invites participation of others. This can be the case even in simple actions upon physical objects (e.g. an attempt to push a fallen leaf away from your face) but is especially so for interpersonally directed intentional actions (e.g. an attempt to give a kiss or toy to a baby). The action is open because it can change its course depending on the participation of the leaf (which may move again with the wind just as you reach for it) or of the baby (who may burst into tears as you approach) and because when it is directed to animate beings it is provoking participation. (Reddy 2008, 162)

On this way of conceiving them, intentions are interactive phenomena, and testing for intentionality is a way of interacting (and indeed seeing whether one can interact) with another. If the movement of something is complete in a way in which its course is determined whatever one does, then it is difficult (if not impossible) to interact with it; this is why the understanding of intentions as 'incomplete' is part and parcel of an understanding of interaction as 'open rather than scripted' (Reddy 2008, 176). Immediately after the above passage Reddy gives another example, which allows us to avoid a misreading:

> Sometimes it is simple muscle tension that conveys this openness or incompleteness and invitation to participate. There is a little clay sculpture on my kitchen window: a woman with her body twisted sideways, one hand gripping her own shoulder, her face rapt, listening, the sculptor told me, to an invisible flute player. Watching her, I can feel her striving in my own body. The key to awareness of intentions may indeed be the experience of being moved by movement. (Reddy 2008, 162)

The case of feeling the statue's intentions may seem to call for a different explanation (to the above two examples) – one could, for instance, explain it as an impulse to do the same thing, to imitate, in this case by completing the action, perhaps as a result of the internalisation of certain scripts (e.g. having learnt to 'strive', or to 'listen to flute players' in some specific way), but this would be a misreading of the case. It is, instead, better understood as an echo of our responding actively to – of relating in an active way with – others. The point is not to do the 'same' thing, but how easily we are provoked to action (action that we experience emotionally). The emotion is not a carrier for re-enactment of some learned action, but rather the medium via which we show how we can be affected by how things appear to us, how we can be moved by others; the other's intention is experienced emotionally as an invitation, an opportunity to interact – the point is not to complete it in some specific way, for it could be continued in all kinds of ways; indeed, the point is precisely in being able to continue in all kinds of ways one cannot quite predict or control. Of course, these are snippets of action or interaction: the point is to show glimpses of both the openness of interaction (which we can see partly by showing the incompleteness, or ongoing-ness, of intentions) and playfulness (an echo of which we witness when we complete – or feel we are completing, or carrying on – what we experience as another's intentions).

We mentioned humour above, and noted how, according to Reddy, it required a mix of engagement and disengagement (and how it might help us see disengagement as part of, and not outside, interaction). In a chapter dedicated to humour (chapter 9), Reddy that humour is 'a fundamentally relational thing' (Reddy 2008, 183), and, further, a relational thing that incorporates – is enhanced by – that mix of engagement and disengagement. By calling humour relational, Reddy is able to focus less on 'the structure of joke' – thereby defining funniness exclusively by virtue of some property (e.g. incongruity or violation of norms) of the joke – and emphasise instead how 'the roots of humour lie…in the motivational / emotional

process of being interested in others' emotional reactions, in particular in this case, laughter' (Reddy 2008, 206).

Reddy offers many examples of interaction between infants and adults where, she argues, it is the interaction – including the very fact that the other person laughs – that is the source of funniness (see, e.g. Reddy 2008, 201-2). We laugh more in company with others (there are cases of solitary laughter, but they are rare), and laughing less is more likely to be a matter of not responding to (not finding interesting) others' laughter and not a matter of not understanding or not recognising a property of the joke according to which it is allegedly funny (in this respect, the data on autistic children, and their lack of interest in others' laughter, is particularly interesting: see Reddy 2008, 203-4).

When Reddy comes to discuss disengagement, she faces a difficulty, for insofar as she thinks of humour as relational, and she thinks of relating as perfect engagement (without disengagement), the very phenomenon of disengagement may appear 'paradoxical' (Reddy 2008, 209; put another way, if she defines the source of funniness in engagement alone, and not also in disengagement, she makes it difficult for herself to account for how disengagement can be a source of funniness). But what is interesting is that in her examples of disengagement – e.g. teasing – Reddy shows how the relation with the other contains both drawing in and withdrawing, a coming in and out of the relation that in fact is part of, and deepens, the interaction. What she calls 'disengagement' in this context is more like an explicit signalling – a pointing to the very interaction itself, making the relation artificial in some way – that makes the interaction more meaningful to the persons participating in it. In this context, Reddy describes (see 2008, 210-211) fascinating reports of infants in the first year of study crawling to one of their parents and then withdrawing, or pretending to get hurt, or interrupting someone else doing something. What is particularly noteworthy about these cases is that the infant looks for a reaction from the parent, and thereby either initiates an interaction (where there was not one, e.g. when someone was doing something, not paying attention to the infant) or deepens the interaction, e.g. where the interaction was (perhaps) becoming too predictable (as in the crawling case). The point is to see how the tease is an invitation to create or intensify a relation with another: the tease does not work in isolation, and only works when the invitation to begin or renew relations is taken up. A similar point – about renewing, deepening communication – is made about forms of deception (faking, pretending) in infants and

children (see Reddy 2008, chapter 10). What the infant initiates here is not a script –
they are not looking for some predictable reaction that they can respond to with an
equally predictable (expected) reply – but in fact the opposite, i.e. something
unpredictable, something unrecognisable.

What we are seeing here is that it is easier to understand this when we pay
attention to the quality of interaction (in this case, between parents and infants), a
kind of quality that depends on maintaining difference between the interactants as
well as openness (unpredictability) in the relation. What Reddy's examples also allow
us to notice is how the quality of interaction (from this second-person perspective)
needs to be matched by playfulness in the interactants, a kind of interest in others'
reactions, and a willingness to create a space for a relation that is not a matter of
enacting something previously agreed on or committed to (or internalised in some
way) but precisely a space where neither side can be in control of what may happen.

B. Play in the Development Literature

Before we leave this brief discussion of Reddy's second-person perspective, it will be
instructive to note how it contrasts with other accounts of infants / children at play.
We have had, in this thesis, several occasions to refer to the difference between games
and play (see, in particular, Section 7.C.). What we see in the literature on play –
including in the developmental literature – is a tendency to characterise play as a
matter of playing games (where games themselves are understood in a specific way,
i.e. a rule-governed way). Where room is made to see children as playing in a way
that does not involve the mastery of and commitment to rules – as for instance so-
called 'practice games (e.g. sensory-motor exercises done without practical purpose)'
and 'symbolic games (e.g. make-believe and pretence)' (Paglieri 2005, 149) – these
tend to be subjected to a hierarchy that favours rules, i.e. playing without rules is said
to be primitive, such that it is best understood as a step better overcome as soon as
possible so that one can (like adults do) participate in games with rules (this was Jean
Piaget's approach, as described recently by Paglieri 2005). In fact, part of the
response to Piaget's schema (from practice to symbolic to rule-based playing) was to
say that even symbolic play was a matter of playing games with rules (this was Lev
Vygotsky's reply to Piaget: see Paglieri 2005, 151). Paglieri's own position is that
children learn to play with rules (not just play games according to rules), in order to
prepare themselves for adult life, which is characterised by the 'predominance of

codified games without rule change' (Paglieri adds: 'participation in games with pre-codified systems of rules is by far the most common playful activity among adults': 2005, 155).

This tendency – to reduce playing to games, or to prioritise and value playing games with rules – is commonly encountered in the development literature: e.g. Devries (1998) uses games with rules as a mode of moral development, where moral development is understood on the Kohlberg (1986) model, which is itself a very rule-based model; or Bekoff (2001), where the idea is that playing is a safe context for learning the 'codes' which determine what is socially appropriate.[214]

The preference for (and thus the developmental value typically attributed to) playing with rules (playing games on a certain conception of games) stretches back quite far, e.g. it is present in Mead (1934) who speaks of acquiring sociability by playing 'organised games' (see particularly Mead 1934, 152-164), and thereby learning to play specific roles that characterise community life. Un-organised or 'free play' is dismissed as something solitary or solipsistic, whereas organised play is thought of highly because it requires taking into account the attitude of the other (see Mead 1934, 154). We have seen how in Reddy it is the opposite: it is playfulness that needs and requires openness to the contingency of the other's responses, not reducing the other to a character in a pre-established script.

One of the most prominent developmental psychologists working on play amongst children today is Hannes Rakoczy, who in a series of papers has argued for the importance of games for development, including moral development. Let us briefly mention some features of his work, showing how it contrasts with Reddy's.

A useful introduction to Rakoczy's views is available his paper, 'Play, Games and the Development of Collective Intentionality' (2007). This paper precedes a number of more recently reported experimental studies (see, e.g. Rakoczy 2008; Wyman et al 2009; and Rakoczy et al 2009), but there is nothing in the later studies that threatens the basic concepts, which are usefully described in the 2007 paper. The first sentence of Rakoczy's paper shows the continuing influence of the demarcation problem (i.e. wanting to show what is unique about human as opposed to animal /

---

[214] One exception in this context, the analysis of which would take us too far outside the scope of this Chapter, is Winnicott 2005 [1971]; for instance, Winnicott speaks of play as 'inherently precarious': see Winnicott 2005 [1971], 70. Winnicott is not discussed in more detail here because his purpose (he is interested in how pretend or as-if play helps the child develop 'objective perception') is a different one to ours, and does not help bring out the distinctiveness of the second-person perspective.

brute life): 'Play is a widespread phenomenon in the animal kingdom. But only humans engage in games, that is, in rule-governed, conventional play activities' (Rakoczy 2007, 53). Rakoczy understands games in a way that will be familiar to us from our analysis in first four Parts of this thesis: he employs the constitutive strategy, arguing that 'rule games are by definition systems of constitutive rules that assign status functions to the objects involved. Chess, for example, is constituted by the system of rules underlying it' (Rakoczy 2007, 57). Rakoczy follows Piaget and Vygotsky in setting out a developmental schema that begins with pretend play and matures into rule-based play; indeed, he sides with Vygotsky in thinking that pretend play, although it 'usually is quite different from rule games in that centres around fictional matters and is normally not an activity associated with fixed underlying rules', nevertheless 'can also be considered a joint activity with (ad hoc) constitutive rules and status function assignment' (Rakoczy 2007, 57). This has the effect of picturing pretend games in a very rigid fashion. Consider the following example:

> Eve takes her cell phone, puts it to her nose, and say, 'Hm, how delicious this banana is.' She then offers it to Adam ('Here, have some'), who starts to make peeling movements on the phone. He then puts the phone to his mouth and makes enthusiastic chewing movements and 'Yum' sounds. Eve joins in, and finally they close the episode by saying, 'All gone, eaten up'.
>
> Although this is not an instance of playing an established game with fixed rules, it is an instance of collectively playing a game with the assignment of transient status functions, making up ad hoc constitutive rules on the spot. 'This cell phone counts as a "banana" in our pretense context' is the central status function assignment. As the scenario unfolds, 'It counts as peeled now' and then 'It counts as eaten up now' enter the scene.
>
> These assignments bring with them a normative structure to the joint activity. 'X counts as Y in context C' means that in C, X ought to be treated accordingly as a Y. In Eve's and Adam's pretense game, once declared a banana, the phone ought to be treated accordingly in the game. Some pretense acts are inferentially licensed in the game, and others are not. Pretending to peel the phone / banana, eat it, or make it into a milkshake are licensed; pretending to drive it or pretending to fax it are not.
> (Rakoczy 2007, 57)

The example, and the analysis of it, works only if one presupposes that the two children, Eve and Adam, share an understanding of what can be done with a banana. In other words, the example is set up in such a way as to make it seem that all that Eve and Adam can do, once they decide that such-and-such an object is a banana, is to re-enact what is appropriate ('inferentially licensed') to be treated as a banana. If

one does not presuppose such shared knowledge, one can see the game as a way of exploring what might be done with a banana, e.g. could it be driven? Could it be faxed? The point here is that Rakoczy assumes that there is something fixed in relation to how a banana can be treated, and that the children already know this, such that pretend games are simply about re-enacting some 'scenario'. This fails to explain how children might have acquired this allegedly shared knowledge in the first place, or indeed why we should think that bananas are the kind of thing that have fixed uses. But there is another problem in the background here: Rakoczy appears to assume that Eve and Adam are enacting (and thus reinforcing) the properties that characterise a banana, but here we can ask: are they not playing with a phone? Could they be seen, then, as exploring the ways in which a phone may be used (rather than re-enacting a banana)? The artificiality of the example, in this respect, is seen in Eve flagging the 'banana scenario' by calling it a banana; this kind of flagging is too obvious to be playful (again, this becomes easier to see when one sees that the point is to explore the different – precisely unpredictable – ways a phone can be treated).

From Reddy's second-person perspective, all this – even with its emphasis on exploration – would presumably still be too rigid. It is only from the perspective of a presupposed normative structure (where bananas are said to have certain fixed uses) that we describe the point of the interaction as one of re-enacting and reinforcing such a structure; when we, instead, begin from the second-person perspective, and focus on the relation between Eve and Adam, what we might see is mutual provocation – looking for a response from the other, looking for the other to react. Further, and arguably again from Reddy's point of view, the quality of the interaction is enhanced not by Eve enacting out what Adam might expect, but precisely because Adam does not know where Eve might take his provocation. Rakoczy assumes there is an 'episode' that is 'closed up' when Eve says 'All gone, eaten up', but to describe this as closure is possible only when one imposes a certain structure of intelligibility (from the theorist's perspective) onto the interactional scene. What one misses as one does so is the importance of the interaction continuing, which cannot happen if two persons are merely playing out a scenario, but depends precisely on them teasing and provoking each other in unpredictable ways (e.g. by going on to pretend to chew her hand, thereby looking for a reaction from the other). What we see, then, is that Rakoczy misses out on Reddy's sense of playfulness precisely because Rakoczy, unlike Reddy, begins with a structure of intelligibility (a scenario, based on assumed

ways in which a banana can be treated) imposed on the interaction (such that it is described as a predictable enactment or re-enactment, further reinforcing a knowledge that is in fact assumed already to be shared).

Notice that the above example, as described and analysed by Rakoczy, may work well when one pictures Eve to be an experimenter who introduces, imposes, and then closes off a certain structure of intelligibility (a certain scenario), but then what we would see here is that rather than an interaction from the second-person perspective, this would be a case of Eve being in control of the interaction in such a way that Adam is but someone who plays 'her game', and not someone who can change the course of the interaction. In other words, what we would see in such a case is not Eve's playfulness, but rather her rigid commitment to some structure, and her treatment of others not as different, but like her (sharing the same commitment). What we would be miss, if this would be our central case, is that in a playground, Eve's responsiveness to contingency and change (her playfulness) would be tested, e.g. rather than simply playing out a scenario (as the well-behaved Adam does, sticking to the assumed script), Adam may respond at first by peeling the banana but then, once peeled, to sing into it, looking to see how Eve will respond: the point is that the interaction dies off when one thinks of it as merely re-enacting scripts imposed by one party on another. Indeed, to play, Eve must learn not to be rigidly committed to a scenario she introduces, expecting everyone else to conform to what she wants, but instead to be open to being moved in another direction, and 'to run with it': it is her capacity to run with it that will be tested in the playground, for it is that capacity that is needed for her to participate in playing (in the kind of open interactions that characterise playing). What Eve will find, if she only plays with others when they re-enact the scenarios she has in mind, is that she will be very lonely; learning to play is not about imposing some structure one then expects everyone else to conform to, but about developing abilities to invite others, to offer them opportunities for responding, and to respond to them in turn, and thereby to initiate, maintain, renew interactions (as understood from the second-person perspective).

We have looked at an example of pretend play because it is in that context that Rakoczy is prepared to loosen the sense in which rules allegedly govern the playing. In the case of games imagined to be constituted by rules, what we saw above as rigid commitment and the re-enactment of a scenario or structure that is assumed to be shared becomes even more pronounced. The experiments that Rakoczy uses in this

context are problematic (because self-confirming): they depend on teaching the child (and flagging to the child that one is teaching them) a 'game', where there are certain specific ways to do things that count as playing the game. An experimenter shows the child novel objects and declares 'I'll show you a game, it's called "daxing" [novel verb]', and then proceeds to make moves (including moves that the experimenter flags as incorrect by saying, 'oops! That's not daxing') (see Rakoczy 2007, 61-2). After the child is shown the game, and allowed to 'play' with the objects themselves, a puppet is introduced by the experimenter who 'violates' the 'rules' (the ones just flagged by the experimenter as 'not daxing'). Rakoczy relies on these studies a great deal, arguing that they show 'young children's (three year olds') understanding of the context-relative normativity of constitutive rules)' (Rakoczy 2007, 62).

What is problematic about these experiments from the perspective of investigating play is that a certain structure is introduced to the child, and the child then plays by himself, such that the only contingency introduces involves a puppet (and not another child) 'interrupting' the child's 'playing' with the objects. What we see here is the absence of interaction, and a methodology that presupposes, and then looks for commitment to (hence the self-confirmation), some rule-defined structure. For instance, what would be interesting in the above experiments, would be to see how the child would respond to the experimenter changing, after having flagged as inappropriate, some way of treating the objects, but in the course of interacting with the child (not teaching them) – this would add a second-person element to what is otherwise an experiment virtually absent of interaction (notice also the passive position in which the child is positioned in the teaching phase of the experiment: the child is told what is appropriate). In effect, what we see here is how a third-person methodology (one that imposes a certain structure, backed up by a very rigid understanding of the nature of games), leads to third-person results (i.e. ones that make it look as if persons care more about, and are guided by and motivated by commitment to, rule-defined structures than other persons). The point here is not to question the value of these experiments – e.g. they could still be used as ways of understanding how children react when faced with instructions presented as fixed structures by someone with authority (the experimenter) – but to question their status as informing our understanding of play and playfulness.

This thesis is not the occasion for a full account of play, which has a rich history not only in the developmental literature, but also within the humanities and

social sciences, including some treatments which problematise the idea that in order to play we must always and already (even in ad hoc fashion) share some commitment to a rule-defined structure.[215] Our point in calling to mind the literature on play is but to show how a certain understanding of normativity (based, in part, on an analogy with games, themselves understood very rigidly) can lead to missing the entire phenomenon (i.e. play and playfulness) that the theorist sets out to study. Reddy's work – precisely because of its emphasis on the second-person perspective – offers many resources (including methodological recommendations for theorists using experimental methods)[216] for resisting the reduction of play and playfulness to (a certain reductive understanding of) games. Let us turn now to Hanneke De Jaegher's view to see how she can further help us understand interaction from the second-person perspective.

---

[215] See, e.g. Jerolmack 2009, who speaks – principally on the basis of considering playing between humans and animals – of interactions 'without the need to assume "minds" or shared meanings': Jerolmack 2009, 376.

[216] Reddy calls for a 'second-person methodology' in developmental psychology, one which involves theorists interacting in a certain way with the subjects they are researching (see Reddy 2008, 32-34). Reddy relies here, in part, on Hobson's proposals (see Hobson 2002, 11-14).

Chapter 21. The Irregular in Social Interaction

From the second-person perspective, we can see that interaction involves a mix of the regular and the irregular. Acknowledging this, therefore, requires us to amend the explanatory object or target of the social, including its normative dimension, that we might otherwise adopt on the governance view. On the governance view, to recall, we set up an object or target for explanation that is ordered – either ordered, from the beginning, with a mix of the merely regular and that which is added on top (an order of, say, conventions and norms) or a normative structure (of rules) that is thereafter portrayed as internalised and regularised. The governance view has no time for the irregular. From the second-person perspective, however, the presence of irregularity in social interaction is vital.

Given the importance of the irregular for the second-person perspective, and the ignorance of its importance by the governance view, this thesis may sometimes risk over-emphasising irregularity. If this done, it is for the sake of making the argument clearer, i.e. to really bring home the importance of the presence of irregularity in social life. To help us in this task, we need to consider, even if all too briefly, the work of Hanneke De Jaegher. Having done so, and having by then a better grasp of the importance of the irregular for social life, we will turn, again briefly, to consider the kinds of experiences persons have when they experience, in the course of relating with others, that irregularity.

A. *The Lack of Perfect Synchrony*

The most complete – though as yet unpublished – version of De Jaegher's view can be found in doctoral thesis, 'Social Interaction Rhythm and Participatory Sense-Making: An Embodied, Interactional Approach to Social Understanding, with Some Implications for Autism' (De Jaegher 2006). De Jaegher has since then collaborated with a number of authors (e.g. De Jaegher and Di Paolo 2007; Fuchs and De Jaegher 2009; De Jaegher and Froese 2009; McGann and De Jaegher 2009; and De Jaegher, Di Paolo and Gallagher 2010), and all these papers do contain aspects of her own view, but they also contain other ideas that are not necessarily congenial to the direction we want to encourage here. In fact, what interests us here, even with respect to De Jaegher's own work, is a small slither of her approach, and one relating

principally to the kind of asymmetry, of falling in and out of balance, that she attributes to the rhythm of social interaction.

De Jaegher first distances her account from those that emphasise either 'one-sided coordination' or 'pre-coordination' (De Jaegher 2006, chapter 6). In one-sided coordination, interaction is studied from the perspective of one person imitating another; of one person doing what another does, such that this forms the basis of the interaction. In pre-coordination, it is said that there is some agreement prior to the interaction and that this enables and structures the interaction. Neither of these approaches, argues De Jaegher, captures the sense in which two parties may need to adapt to each other during the course of the interaction, and also the sense in which the interaction itself can take on its own rhythm and change the course any of the parties expected or anticipated.

In making this argument, De Jaegher draws, in part, on Alan Fogel's (1993) work, where Fogel spoke of 'co-regulation', which De Jaegher describes as 'a continuous process of mutual adaptation' (De Jaegher 2006, 105). When the two parties interact, they need not have an ultimate end in mind, at which they both aim; nor need it be the case that one party will impose an end and / or means to that end, such that the other must conform. Again, whatever does ensue in the course of the interaction is not agreed upon or somehow predicted / predictable / controllable in advance. One of the interesting features of Fogel's account, on De Jaegher's reconstruction, is that 'Fogel assumes meaning is created continuously as part of the interaction, rather than as pre-existing or inherently present in objects or actions' (De Jaegher 2006, 106). De Jaegher describes one of Fogel's examples in the following way:

> He gives the example of someone at a party who is thinking that she would like to leave soon. This intention is often only partially conveyed, by finishing a drink, taking one's coat or moving gently in the direction of the door. These actions can also go in other directions still though. The person doing them is still open to possibilities, someone might grab her still, a conversation may still ensue. A person's actions and their meaning are created and changed in and through the interaction. For instance, only when the host opens the door for her, will the meaning of 'leaving' be fully created. (De Jaegher 2006, 106)

What is interesting here is that the case is being described as one where neither of the parties (let alone the observer) knows what might happen. In Fogel's own words: 'a

continuous process of mutual social co-ordination requires that there be a continuous unfolding of individual action that is susceptible to being continuously modified by the continuously changing actions of the partner' (quoted in De Jaegher 2006, 107). Notice, for instance, that there is no reference here to both parties sharing some shared framework – some script, say – that they then both re-enact. It is important to see that this is in part enabled by the very way the example is described.

We have been quoting Fogel as depicted by De Jaegher because what is interesting is what De Jaegher does next. According to De Jaegher, Fogel focuses too much on 'good' interactions, i.e. interactions that are fluid and smooth, and where, as a result, 'the correspondence between the interaction partners…is too close' (De Jaegher 2006, 108). In distinguishing her approach in this respect from Fogel, De Jaegher also draws on Scott Kelso (1995) who, she says, 'explains that coordination does not always have to be perfect, that is, that coordination can be relative' (De Jaegher 2006, 110). Kelso's famous example is that of a child and an adult walking together, holding hands: 'one or the other has to adjust either the frequency of their step or the length of their stride', i.e. there is no 'pure phase locking', but something 'more variable, plastic' – there must, in short, be imperfect synchrony (De Jaegher 2006, 110). 'Systems in relative coordination', De Jaegher adds, 'do not phase lock or entrain perfectly. Instead they show phase attraction, which means that they go near perfect synchrony, and move into and out of the zone that surrounds it' (De Jaegher 2006, 111). De Jaegher's conclusion is that what we must pay attention to when studying interaction is the 'regular irregularity of real-world interactions' (De Jaegher 2006, 113), and she summarises as follows:

> Interaction and coordination are complicated events that take place in complicated and variable environments, on the basis of complicated, imperfect mechanisms, with forces pulling and pushing from all sides, which all it make it unlikely that the process will be smooth. In order for individuals to be 'successfully' involved in such complex phenomena, there needs to be a degree of flexibility in the individuals, and, arguably, also in the process. (De Jaegher 2006, 113)

Further adding a moment later:

> …coordination and interaction are not perfect. In order for them to be able to generate and transform meaning, they cannot be smooth. If coordination and interaction were perfectly smooth, as described slightly caricaturally [sic] in

> the discussion of Fogel's work, they would lock interactors [sic] together perfectly, and there would be no space for the emergence of meaning. The flexibility in interaction is not only needed for physical reasons, it is also needed for reasons of meaning generation. Breakdown of coordination, imperfect coordination and repair…are themselves conducive to social understanding. (De Jaegher 2006, 114)

There is much of importance to take from these passages, but we have to be careful. For instance, we do not wish to suggest that interaction must be studied from the perspective, or under the rubric, of coordination, no matter how relative (as opposed to absolute) this is characterised. Notice, for instance, that were we to do so, adopting for instance phrases such as 'the breakdown of coordination', we would be back to where we started, namely with a view where some kind of perfect fit or synchrony is the aim, and anything else needs to be repaired in order to return to equilibrium. Instead, what we ought to take from De Jaegher is that interaction itself is irregular (in the sense of being a mix of balance and imbalance); and further, not only is there no need for perfect synchrony, but 'the generation of meaning' as she puts it is dependent on the absence of such perfection: on a looseness in the process, something incomplete, disproportional, askew and aberrant (language fails us again, because it makes it seem like these are negative notions to be avoided).

Notice that the irregularity De Jaegher points to is part of both the means and end (she does not speak in these terms, and we would prefer to avoid them too, but referring to them helps to avoid a possible misunderstanding): the idea is not that we can take all kinds of paths to the same end (to actually walking together, or to actually co-ordinating in some way); instead, the idea is that both the means and the end are loose and irregular (what we might take to be the means change because what we might take to be the end changes, in the course of the interaction, in unpredictable ways). In other words, what we must avoid is imposing a particular scheme of what it means to walk together, and then depicting the behaviour of persons coordinating to do just that, for 'what it means to walk together' is negotiable by the 'interactors', and not controllable / predictable by either party. For instance, there is no kind of walk – no pre-given structure, no script – that both father and daughter have to conform to; 'what kind of walk' they are on can shift and vary as they go along, e.g. as the daughter points and asks questions about what is to be seen in the surroundings, and as the father speeds up a little looking at the time (the two parties, in order to interact, must adapt to each other – must be capable of being influenced by each other; De

Jaegher calls this 'rhythmic capacity', which we will say more about in a moment). It is this looseness – allowed partly by the description of the example, and partly by seeing how the parties to the interaction need to be flexible and responsive to each other – that we need to keep an eye on.

None of the above, it should be added, is at odds with the idea that we may sometimes draw on certain normative resources to guide ourselves, or to evaluate ourselves and others. What is important is that one does not – in describing the example – refer to normative resources as if they had a metaphysical status, i.e. as if they were capable of organising the environment for the parties in the interaction, such that the possibilities for how those parties interacted would be determined by those ways in which the environment was organised (going on to describe persons as learning to internalise, and then re-enacting, such determinants). If we are walking together, either or both of us may have certain normative resources in mind about how we ought to do so, and these may help us, and we may use them to regulate our conduct, but we may also not, and even if we do, these do not have the power of determining what we do; our bodies, when walking together, are adapting to each other in a loose way – indeed, it is precisely this looseness that allows us to achieve the imperfect synchrony that is needed for interaction.

Chapter 7 of De Jaegher's thesis is dedicated to further elaborating on the 'interaction rhythm' and we must take a brief look at this. Again, what is of interest about De Jaegher's treatment is that she emphasises the relativity and variability of rhythm in an interaction: continual, perfect synchrony is neither necessary, nor desirable for interaction. This does not mean that persons interacting will not, at times, tend towards synchronisation – they may imitate each other, for example, and this can happen without them realising it[217] – and it does not mean that two persons who interact with each other often do not establish certain familiar ways of relating to each other, enabling them to anticipate each other relatively successfully. But these are all contingent, and also inherently fragile, possibilities: we need not think that interaction, when it 'works', results in and maintains perfect synchrony, or that our anticipations of each other are necessarily structured either by us sharing the same cultural script or some 'interactional history' or 'interactional relationship' (see De

---

[217] An interesting example De Jaegher mentions in this context is the synchronisation of heartbeats between 'client and therapist in a session of improvised music therapy' (De Jaegher 2006, 133; here, De Jaegher draws on the work of Neugebauer and Aldridge 1998). As noted above, however, by reference to Reddy's work, imitation may itself need to be understood in second-personal terms.

Jaegher 2006, 147). Our basic ability to interact with one another is a loose and tentative one in which we are responsive to the contingencies of each other.

Let us say few more words about this notion of anticipation. Anticipation can be exercised rigidly, as when one anticipates (or, more strongly, expects, including possibly normatively expects) someone to do something, such that (if the expectation is a normative one) one is committed to seeing anything else the other may do as a mistake (for one sees it as a violation of the expectation), and *sometimes* even as a mistake to be punished.[218] Anticipation can also be exercised more flexibly, even in the sense of a normative expectation, as one thinks that there is a range of possibilities, or some rough aim, the other would be recommended to take, but one leaves room for the other to change one's view about the appropriateness of the means or the end. But even anticipation understood in this more flexible way (and certainly anticipation exercised in the rigid way) is not enough – i.e. we should not think of interaction, understood from the second-person perspective, second as necessarily, and certainly not exhaustively, structured by two persons stocked with anticipations (especially shared ones). Certainly, we may on occasion anticipate (and we may even sometimes share this anticipation) – in other words, any one relation with another person may be peppered with anticipations, and these may be exercised more or less rigidly, and may sometimes be shared – but it is not the case that interaction is adequately describable on the basis of two persons having to follow, by default, such anticipations unless given reason to do otherwise. We are not, in interacting with each other, always and necessarily anticipating and on the eager lookout for confirmation of our anticipations.

Understanding this may be assisted by how De Jaegher speaks of 'rhythmic capacity', precisely that capacity that allows us to participate in the variability, irregularity and contingency of interaction rhythms. De Jaegher defines rhythmic capacity in general as 'the capacity to flexibly, temporally coordinate, through and in the interaction, with another person' (De Jaegher 2006, 146), though the reference here to coordinate can be misleading for it may suggest that the capacity is one designed to achieve and maintain perfect synchrony. The following passage is, in this respect, clearer:

---

[218] To say that such a potential evaluation is *always* accompanied by punishment would be to miss how, in interacting with each other, we will often do everything to avoid signalling criticism of the other, especially when the other is a non-intimate.

> …better social interactors [sic] are those with a greater rhythmic capacity. A greater rhythmic capacity is one that is more flexible. Worse interaction partners are those with a less flexible rhythmic capacity. Good social skill is defined here as a capacity to deal with many different social partners in different situations, with highly unpredictable partners, partners whose sense- and world-making is very different to one's own… Better or worse interaction partners are characterised by whether, in a certain social situation, they are better or worse at dealing with change and unexpectedness in a social interaction. (De Jaegher 2006, 152)

What is so instructive about this passage is that it helps us see how our experience of another can (and indeed needs to) be ambivalent. We make a mistake when we think that for interaction to go well, we must be so attune to one another that we can immediately and non-inferentially tell 'what' the other is doing and about to do, and we make that mistake partly because we picture us enacting or fitting into a pre-existing mould of what would make an ideal interaction. In other words, when we picture persons as interacting well insofar as they reach and maintain some sort of perfect synchrony (re-enact some script, act in accordance with the rules structuring some interaction, blindly fall into patterns of interaction that happen to be in accord with the rules, etc), then we are also likely to picture persons' abilities to relate to one another as ones that prepare them for that: i.e. we then picture persons as stocked with default categories and scripts, or having developed dispositions that are sensitive to pre-established patterns, and as immediately recognising / following / enacting these blindly until and unless they are given reason to change them. The point being made here – building on the above aspect of De Jaegher's analysis – is that this pictures both interaction and our abilities much too rigidly and narrowly. Interaction going well does not mean it slips into some pre-determined mould; on the contrary, interaction going well means it is somewhat irregular, unfinished, incomplete, and somewhat imbalanced. As soon as we see this, we also see that our ability (including our ability to anticipate) must be more inherently flexible, more capable of hesitation, waiting, not over-committing itself; in a way, our ability must itself be unfinished, irregular, incomplete and somewhat imbalanced: the ambivalence of our capacities, is their strength, not a weakness to be ironed out.

Again, one vital thing it is important to see in all of this is that we are tempted – but must resist the temptation – to picture forms, moulds, ideas, structures, patterns etc of interaction that we identify as default / ideal, such that we imagine that when

things go well, they go so smoothly as to simply fall in line with, accord with, such forms, moulds, etc. This also means resisting an understanding of change, tension, or breakdown as something negative or external to an interaction going well; change must not be thought of as something that occurs only when something happens that disrupts the status quo, how things really are and should be, unless we are given reason to see them / evaluate them differently. Instead, change, tension, and breakdown are internal to both interaction and our abilities to interact. This is hard to conceive, for we are used to thinking of these concepts as negative responses to things not going as we expected / predicted / hoped they would go. That is why we need descriptions of examples that allow us to see that things going well mean that it is not pre-determined how they will go.

Perhaps the difficulties here are exacerbated by the way the mind takes a certain pleasure in confirming itself, i.e. it ascribes a certain status to those ideas it might self-consciously use to justify or systematise things in a certain way. In doing so, the mind grants a dual / Janus-like face to ideas: both as default descriptions of how things are and as ideals (as depicting how they should be), with the result that we think that the way things should be (on any one scheme of justification or systematisation) is the way they are unless we are given reason to think otherwise. It is as if the normal state of affairs is for our schemes to apply, as if they were to apply themselves. But what we notice when we pay more attention to describing the examples (of interaction, say) carefully is that neither we, nor any interaction, is by default conforming to some particular pattern or rule-given structure unless and until it is changed; what happens here is that we translate the experience (we can have) of treating things as happening in accord with how we can conceive of them (of treating things as unchangeable, as ideal) into a default status where they acquire metaphysical powers (they apply themselves to dictate how things are). It is this that we must resist if we are to make room for an alternative to the governance view. Hopefully, thinking about interaction rhythm and rhythmic capacity (as De Jaegher discusses them, at least in part) can help us do that. It is, of course, but a step on the way, and we will be returning to the issues raised in this Section below (e.g. especially with regard to how to conceive of 'breakdown', in Chapter 22).

B. Experiencing Irregularity

We cannot hope, in the confines of this thesis, to capture all the important facets of experiencing others in the course of interacting with them, when that interaction is understood from the second-person perspective. For instance, we have not here included discussion of the I-You iterations that might, in fact, be of great importance to interaction understood from the second-person perspective, i.e. iterations of the form: 'I see you seeing me seeing you', etc. In other words, for a fuller description of interaction from a second-person perspective we may need to delve deeper into the mutuality of consciousness of persons in such relations.[219] For present purposes, what is vital is that we make the second-person perspective on interaction plausible, and also point to some of what it enables to see, especially with a view to describing (in Part VI) the abilities that persons need in order to interact with others more or less successfully. What is being stressed about interaction so viewed in this Chapter is the importance, within that interaction, of irregularity. In this Section, we want to give some further indication of the variety of experiences persons can have of that irregularity.

In his recent book, *Together: the Rituals, Pleasures and Politics of Cooperation* (2012), Richard Sennett provides some striking examples and descriptions of experiences that persons have in the course of interactions that are, as he puts it, 'dialogical', or better, dialogically conceived. For instance, Sennett points to the 'dialogics' present in musical rehearsal (speaking from his own personal experience). You might think, he notes, that musical rehearsals aim at homogeneity or perfect harmony; but this, he asserts, is not the case. Instead, 'musical character appears…through little dramas of deference and assertion; in chamber music, particularly, we need to hear individuals speaking in different voices which sometimes conflict, as in bowings or string colour. Weaving together these differences is like conducting a rich conversation' (Sennett 2012, 14-15). The experience of these 'little dramas' is an example of the experiences of irregularity in interaction from the second-person perspective.

---

[219] Interestingly, one theoretical context in which such mutuality has been explored is in the literature on joint attention where, for instance, it is said that not only do we need to be both focusing on the same object, but also be seeing that the other sees we are seeing 1) the object; and 2) the other is seeing the object; and 3) the other is seeing us see the object. For a recent collection exploring these themes, see Seemann 2011.

In his book, Sennett focuses on the skill (or craft)[220] of mutual responsiveness. At times, and as also indicated by the reference in his title to 'cooperation', Sennett seems to have in mind the attainment of a certain ideal togetherness – a kind of fluency in relating with each other; perhaps a kind of embodied order. To the extent that his discussion is set up in this way, his examples and descriptions are of less value to the second-person perspective. However, as is indicated by the above example of musical rehearsal, a good part of his discussion is of great importance to our understanding of the variety of experiences persons can have of irregularity in social interaction conceived from the second-person perspective.

The characterisation of musical rehearsal above as 'dialogic' is important. Certainly, and as noted in the Introduction to this thesis, the literature on 'dialogics', including the dialogical analysis of communication, is a very important one for any future development of the second-person perspective. When Sennett uses the term, he says that by it he means roughly what the Russian literary critic, Mikhail Bakhtin, meant by it, namely: 'a discussion which does not resolve itself by finding common ground. Though no shared agreements may be reached, through the process of exchange people may become more aware of their own views and expand their understanding of one another' (Sennett 2012, 19). We see here how connected the dialogic is to the contingency – rather than necessity – of sharing, and thus why it was important to distance ourselves from that approach to the social, including its normative dimension (as we did in Part IV of this thesis).

Where Sennett adds something of value to our discussion is where he speaks, for instance, of the 'mood' required for a 'dialogical conversation'. The mood in question is the 'subjunctive mood' – one characterised, we might say, by tentativeness, by the presence of a 'perhaps' or a 'possibly'. In Sennett's words, 'the subjunctive mood is most at home in the dialogical domain, that world of talk which makes an open social space, where discussion can take an unforeseen direction… By practicising indirection, speaking to one another in the subjunctive mood, we can experience a certain kind of sociable pleasure: being with other people, focusing on and learning about them, without forcing ourselves in the mould of being like them' (Sennett 2012, 23), or, indeed, we might add, without forcing them to be like us. Of course, we have to be careful here and avoid introducing any dichotomy between

---

[220] Sennett's book follows on from his earlier *The Craftsman* (2008), which was a more general defence of craftsmanship.

'open' and 'closed' social spaces – the point being made in this thesis is that any social space includes both open and closed elements; it is a balance or mix of the two. The danger, to repeat, with analysing the normative dimension of social life is that we will depict the social as, at least at certain times, completely closed, as forming a closed order to be explained (by the theorist) and participated in (by the social agent). With that proviso, we see that what Sennett enables us to notice here is the presence of a certain kind of tone or flavour to the experience of interaction from the second-person perspective. We also see, once more, the importance of experiencing the other as different, where that difference is not seen as an obstacle to be overcome or a wrinkle to be smoothed out.

The indirectness that accompanies the subjunctive mood points also to other kinds of experiences: irony, for instance. Experientially, one might further cash this out in terms of waiting, hesitating, or being alert without asserting any one particular way in which the interaction ought to go. In a fascinating discussion, Sennett finds precisely such ironic forms of relating in the practice of diplomacy. Referring to Castiglione's sixteenth-century book of advice, *The Book of the Courtier*, Sennett notes that Castiglione placed much value on '*sprezzatura*', literally meaning something 'springy', or, as Sennett puts it in the context of diplomacy, 'lightness in gesture and speech' (see Sennett 2012, 211). This medieval Italian concept is by no means forgotten today: for example, Sennett shows how Sir Ernest Satow's *Satow's Diplomatic Practice* (the authoritative text for professional diplomats, first published in 1917 and now in its sixth edition) also emphasises similar 'informality, indirection and mutuality' (Sennett 2012, 238). For example, Satow advises the occasional use of a '*bout de papier*', a technique that involves passing an unsigned piece of paper across the conference table which contains versions of the formula, 'If you felt able to propose…I should be prepared to try it on my government', such that 'the diplomat behaves as if he or she is then responding to an adversary's position rather than asserting one of his or her own'; this creates, as Sennett notes 'a space of deference from a position of power; it is an exercise in applying minimum force' (Sennett 2012, 238). The exchange is 'liminal in the sense that it creates ambiguity' (Sennett 2012, 239). The presence of these rituals of indirectness – or, precisely of irony – involving deference, ambiguity, non-assertiveness, are vital for diplomatic interaction, an often an indicator of such interaction going well.

It is important to add here these experiences of ambiguity and hesitation are not merely facets of preparing or initiating an interaction understood from a second-person perspective – they are also experiences that occur during interaction, so conceived, itself. They are, indeed, experiences of the irregularity of social interaction. In the case of the '*bout de papier*' technique, there are moments before, during and after that may feel awkward, where one is uncertain about both the act of using that technique or how the other will reply. There is some anxiety here; some risk; some stepping on each other's toes. Not everything is fluent or synchronous or harmonious. There is some meandering, some tentativeness. There are fits and starts, including false starts and the promise of ends that do not, and need not, materialise. There are unresolved tensions. All these, and no doubt other, experiences are part of an interaction from the second-person perspective. They need not dominate it – they can accompany, for instance, contingent reference to certain normative resources, including even the treating of some rule, for example, as necessary – but they need to be there, at least in part. The importance of their presence should already give us an indication of how the breakdown of interaction is conceived on the second-person perspective. To elaborate on this further, we need to turn to the next Chapter.

Chapter 22. Interactional Breakdown

It matters a great deal how 'interactional breakdown' is conceived. What we will try to make room for in this Chapter is an alternative to the usual understanding in which theorists posit smooth, easy, relations between persons that are said to depend on the regular (blind, internalised) following of necessarily shared determinants of what is appropriate and inappropriate, such that breakdown is then said to be caused by deviations from or violations of such determinants. The alternative involves seeing interaction from the second-person perspective (in the way in which this has been developed above by reference, in large part, to Reddy and De Jaegher's views), namely as a process that involves both a falling in and out of balance, a mutual responsiveness to contingency, and thus a form of relational dynamics that does not depend on an underlying sameness or joint commitment to and application of rules that belong precisely to the rule-governed structure in which participants are imagined / described to be participating in.

There is, then, a certain quality of interaction that is not dependent on repetition or re-enactment of the same, but rather consists precisely in the respect and maintenance of difference, which includes the readiness to wait, be hesitant, to explore how the other might be feeling or what they might be doing, and be susceptible to changing course as a result of the invitations and opportunities provided by the other. In other words, what we see in breakdowns so conceived is the danger of being over-committed to, and over-emphasising the value of, certain determinants of appropriateness and inappropriateness (precisely treating them as something more than normative resources that are subject to the primacy of interaction understood from the second-person perspective).

In developing this alternative, we will be focusing mainly on recent studies of pathologies of interaction between mothers and infants, but we will also mention, towards the end of this Chapter, how the results of these may be related to studies of social interaction from other angles, including the still-face experiments referred to in Chapter 20, autism and prosopagnosia. It has to be emphasised that, in drawing on this psychological and empirical literature, this thesis does not claim to be making a contribution to that literature, or even passing judgement on the validity of any of the results of experiments reported in the literature being discussed. Instead, the point of referring to this literature is to illustrate a conceptual point about how the breakdown

of interaction can be conceived – it is, in short, to make possible and *prima facie* plausible this conception of the breakdown of interaction, because the possibility and *prima facie* plausibility of this conception is important for the second-person perspective on social life.

A. Depressed Mother and Infant Interaction

The studies of mother-infant interaction that we shall briefly consider here stem from research into maternal depression, i.e. where mother-infant interaction is studied from the perspective of diagnosing and curing maternal depression. What makes the studies we shall consider unique – they are all primarily based at a special unit at the University of Heidelberg Medical School – is the use of videos as part of the treatment (known as psychotherapy). The reason why these videos are of interest here is that they allow us to look at when interaction is going well and when it breaks down – but breaks down in a very specific way, as we shall see in a moment. We shall be looking at some segments of one interaction, but first we need to give some background.

For a long time it was assumed, in studies of mother-infant interaction, that an interaction going well was one that was characterised by either perfect or next-to-perfect synchrony. The emphasis was on sameness, including both sameness of action and feeling: 'The maternal ability to physically mimic the child's affective condition and thus to create a "matching" of the mutual emotional expression has long been thought to play a critical role in development' (Reck et al 2004, 274). In other words, it was thought that there needed to be 'synchronicity and reciprocity of maternal and child affective states, along with a high proportion of positive expressions of affect' (Reck et al 2004, 274). In 1997, however, E.Z. Tronick and M.K. Weinberg challenged this long-held assumption, and 'contested the presumed central role of synchronicity' (Reck et al 2004, 274; see Tronick and Weinberg 1997). Here is a summary of their findings:

> They showed, instead, that healthy mothers and their children characteristically interact with only moderate levels of synchronicity and with a modest proportion of positive as well as negative expressions of affect. In a regular play situation, infants and mothers matched their affects only 30% of the time. Tronick and Cohn [(1989)] characterise the typical mother-child interaction as a flexible process, in which there are frequent shifts between affective 'matches' and 'mismatches' ('interactive errors'). 'Mismatches' are

typically associated with a negative, 'matches' with a positive expression of affect. The interactional transition from an uncoordinated to a coordinated state is called 'interactive repair'. In face-to-face interactions with a 6-month-old infant, there is normally a quick repair of interaction mistakes. When a pair experiences an interactive error, they are able to repair it and get back to a matched state within an average of 3-5s… [Later studies] confirmed the hypothesis that normal interactions can be described as a process of matching, mismatching and repair. (Reck et al 2004, 274)

There are two things to say about the above passage: first, notice the importance of a *lack* of matching, and the emphasis on flexibility; second, notice that the language in which the interaction is described is still one of matching (as positive) and mismatching (as negative), where the mismatching needs 'repair' work in order to again achieve matching. The point that needs to be made here is that the data in fact show that the language used above is inadequate and misleading: it is not that matching is positive and mismatching is negative, or that there is an ideal state (of matching) which is the reason why mismatching must be repaired; on the contrary, an interaction going well is one where parties are not necessarily co-ordinating perfectly, but rather responding and reacting to each other's invitations and opportunities to interact – there is, in other words, a responsiveness to each other, but this is not a responsiveness based on doing the same thing (either by mimicking or by re-enacting a common script), but precisely on respecting and maintaining difference. Certainly, there may at times be overlaps in action and feeling, and this can be experienced as a positive thing, but such overlaps are neither the 'norm', nor the 'ideal', and need not be the aim of interaction. What is important, instead, is something that we might call sensitivity to interactional contingency – the creation of a space for interaction that neither party controls or dominates but to which both contribute.

We say 'sensitivity' here (the other term often used is 'responsiveness'), but this can be misleading, as it suggests immediacy, ease and fluency and also stability of what is the object of sensitivity (or responsiveness): on the contrary, what we see in mother-infant interactions going well is relations that are imbalanced, rough, messy, zigzagging, prodding, provoking, mistimed, incomplete, and sketchy with a predominantly non-synchronous, non-monotone, non-harmonious tempo. One could argue this demands a kind of sensitivity (or responsiveness) but this would have to be one that is willing to suspend its eagerness, not over-commit, waiting actively, lingering with interest and curiosity, sometimes leading and sometimes withdrawing,

always opening up and being open to indeterminate possibilities for interaction. This reading is further enhanced when we look at what tends to go badly in interactions between depressed mothers and infants.

When we think of depression, we tend to think of withdrawal, passivity, and inexpressiveness. This image of depression also found its way into the literature on depressed mothers, but this created a much too one-sided and skewed view of what goes wrong in mother-infant interaction. Following on from their studies of interactions going well, Weinberg and Tronick (1998) argued that depression can manifest itself in a variety of ways, including in behaviour that is 'withdrawn, reserved and less engaged', but also in 'intrusive behaviour', where the mother expresses annoyance with and interrupts the child's activities, often using a loud voice, and engages, too readily and too eagerly, in 'positive and dedicated behaviour' (Reck et al 2004, 277). According to them, what was of most importance in these breakdowns was 'disruption of the pairs' capacity for interactive repair' (Reck et al 2004, 277).[221] The best way to see this is too look at an example of an interaction going badly. We shall do this by considering the descriptions of an interaction between a mother called Karen and an infant called Diana, as reported by Downing et al (2008).

The video was filmed in Heidelberg, at the special unit mentioned above. Here is a description of the method used:

> When a mother arrives with her baby in the Heidelberg clinic, an interaction video is filmed the first or second day. This video will be used both for an intervention and at a later point for research. The first video is always of free play. The infant is strapped into a plastic seat, and the mother is seated directly facing her child. The mother is instructed to 'Play with your baby the way you normally would do'. Two cameras are employed. A split-screen image is produced, with separate frontal views of infant and mother. The person responsible for the filming is behind a one-way mirror. (Downing et al 2008, 283)

Downing et al describe two segments of the video totalling 36 seconds (the video in total was 8 minutes long). The first segment is 25 seconds, and the second 11 seconds. Here is the description of the first segment:

---

[221] Recalling what was said above, let us keep in mind that the language here can mislead us, for it is not so much 'repair' that is at issue, but precisely the creation of an indeterminate space for interaction.

For the first 2s, an observer sees only Diana, the 5-month-old infant, sitting in her seat. Then Karen, the mother, appears onscreen and settles into her chair.

She gives a greeting. It is particular. As she lands in her place, grinning widely, her voice goes, 'Heh, *heh!*' The quality is almost explosive. Diana appears unperturbed, however. As she looks at mother, her facial display is neutral, but she seems alert and interested.

From the mother's side then ensues a flurry of laughs, brief statements, and large arm and hand movements. Her hands dart in and out of a space inches in front of Diana's face and chest. Sometimes, she (mother) touches abruptly, but momentarily. Other times she opens one hand, palm towards Diana and fingers skyward, but stops just short of making contact. Her comments have a playful tone, with an ironic, critical edge. Diana's shirt is 'a little dirty, right?' Her face, which apparently also has a speck on it, is 'not okay' ('Das geht nicht') and is promptly cleaned with a hectic swipe.

Most striking of all is mother's own face. It is mobile and vivid, with a rapid flow of expressive displays. These expressions largely manifest themselves as simultaneous opposites, however. Partly, she has a[n] open smile. Partly, she shows continual flashes of anger and also of disgust…

Diana's concurrent behavior during these seconds is just as singular. She appears unfazed. She does look downwards and slightly to her left, focusing apparently on a loose piece of seat belt strap near her own left hip. Otherwise nothing. No cringing, ducking, startling, twisting. She is as if oblivious to the onslaught of sound and gestures directed toward her. Even when her face is wiped hard, one would think she barely notices it.

At the same time, she (Diana) is hardly inactive. On the contrary, the seat belt soon draws not only her gaze but first one hand and then the other. She fiddles with it. Her movements seem concentrated, deft for her age, and unhurried. Her tempo is slow, explorative. She is as if in a different temporal universe, one immune to the fast-moving surrounding events.
(Downing et al 2008, 284)

Downing et al point out that if one took just a quick glance at this interaction, one might not see the ways in which it is problematic. It looks, at first blush, like a 'thoroughly stabilised, workable arrangement', with Karen appearing 'quite at home in the interaction' (Downing et al 2008, 285). When looked at more closely, however, one can see that Karen's behaviour is over-confident, intrusive, abrupt and controlling: 'Nothing looks tentative in her manner, or uncertain… her movements sabotage the contact' (Downing et al 2008, 285). The key point is that 'As things stand, the amount of contingency between mother and infant is quite low', with the effect that 'Diana is trapped in a dilemma. If she synchronised with Karen's speedy rhythm, and Karen likewise with her, the probable result would be excessive "high matching"' (Downing et al 2008, 285), which we have seen is destructive for the quality of interaction.

Now consider the second shorter segment (we do not reproduce it here in full, but just the first initial period, the first few seconds of the 11):

> Here, for a brief interlude, a significantly different exchange occurs. The first step seems to come from mother. Diana is still poking the chair strap on her left side. Mother imitates this after a fashion, or at least mirrors it. She puts her own (right) hand near Diana's (left) hand, her forefinger extended downwards, fingertip touching the seat arm. She lets her finger rest there, about half an inch from Diana's hand. Mother looks down at the two hands. For a moment, her fast movements subside.
>
> It takes less than 1s for Diana to respond. With her left hand (which was poking the strap), she takes a hold of mother's finger. Mother allows this. For about 1 s more, they remain both focused on this two-hand constellation. Diana then slowly brings her right hand over. With this hand, she grasps another part of mother's hand, and then with both her (Diana's) hands begins maneuvering mother's hand. Mother permits this as well. Visually, they both concentrate on the intertwined hands constellation.
>
> Movement continues: a steady, gradual raising or guiding of the large hand upwards. I say *steady* to emphasize the rhythmic quality. Diana seems to be leading, and the temporal style with which she does so is smooth and contained. Mother, following well, has now adopted the same temporality – or in part, anyway. There remains a jumpiness in her facial displays and small head movements. But mostly she organizes her body seemingly to fit the slow arc Diana has initiated.
>
> As the hands constellation rises higher, arriving just below Diana's chin, Diana and mother's mouths simultaneously open, each forming an 'O.' The mutual onset occurs in precisely the same quarter of a second… Diana's mouth then spreads into a radiating smile. Her whole face lights with pleasure. She next edges one side of mother's hand directly into her mouth.
> (Downing et al 2008, 286)

There are a couple of things to notice about this interaction. One could emphasise the period of overlapping actions and (perhaps) feelings in this episode, i.e. one could read as positive the moment when Karen and Diana both form an 'O' together, and one could read as negative what happens afterwards (Karen's breaking of this co-ordination). This, however, would be a misreading of the interaction. The point in this episode (and it is important to keep in mind that this is just several seconds within a 11 second segment) is not merely or mainly the joint action and (perhaps) joint feeling, but in large part the sense in which a space for interaction was opened up where both parties remained different (one leading, the other following, but then switching) in a way that manifests sensitivity to interactional contingency. Recall the moment when Karen looks down at her hands; this enables (does not determine, but offers an opportunity for) Diana to take the initiative, which she does by picking up

Karen's finger; Karen runs with this, and Diana brings her second hand over; matters are proceeding slowly and in an explorative way; there follows a careful tension in the raising of the large hand; and this eventually leads to a co-ordinated movement, the making of the 'O.' It would, then, be misreading this episode to think that the aim and point, and value, of the interaction is the making of the 'O'. This moment is but one in a series of actions that are not co-ordinated (the making of the 'O' need not have happened to have made the interaction one that went well), but precisely sensitive to the contingency of the encounter (of course, none of this is to say that one cannot take pleasure in, and that there cannot be great value in, overlaps of the kind exemplified by the 'O' moment).

Notice, too, that neither Karen nor Diana in this brief spell can be said to be active or passive, or playing some specific role, or enacting some script, or an object for the others' manipulation: both are both actively passive (waiting for the other, reading the movements of the other as invitations) as well as actively active (also taking the initiative, i.e. offering a new movement as an invitation).

Contrast this episode with the longer one above: recall that the movements of the mother there were described as abrupt, as intrusive, and precisely absent of any waiting, any hesitation, any tentativeness or uncertainty. Of course, waiting etc on its own is hardly helpful: the mother could be waiting for the video taping to be over; the point is that in the longer episode there is no interest in the other's difference, initiative or autonomy – the other for the mother is but an object that is part of her plan or scheme. In the episode in which the interaction is not going well (the first episode) the mother ignores the infant's activity (her exploration of the seat-belt for instance); the infant preoccupies herself, while the mother either interrupts or handles the infant (again, as if it were an object). One has the feeling that the mother is committed to doing certain things in a certain manner – correcting the dirtiness of the shirt, for instance – and on making associated 'appropriate' statements and gestures (saying 'not okay'; cleaning with a swipe). What we have here is the imposition of patterns or set ways of doing things and, as a result, precisely the breakdown of interaction.

B. Sensitivity to Interactional Contingency

We have now several times referred to the importance of sensitivity to contingency (in this context, interactional contingency). Earlier, when discussing Reddy's work (in

Chapter 20), we mentioned Murray's still-face experiments and promised to return to them briefly. Murray's experiments were conducted in 1975, and researchers have been analysing and (in part) reproducing them since then. One such recent and important attempt was made by Nadel et al (1999). Let us recall that the point of the experiments is to show how infants 'react negatively to noncontingent maternal behaviour' (Nadel et al 1999, 166). Nadel et al introduced a number of changes that were designed to fend off various other possible interpretations of the data, e.g. they turned off the replay and re-introduced a 'live' interaction (the hypothesis being that the infant would re-engage as soon as an indeterminate interaction was again made available). Indeed, Nadel et al argued that 'the second live session provides the decisive evidence needed to evaluate the hypothesis that infants show an early sensitivity to violations of adult social contingencies during face-to-face interactions', because if the infant was engaged in the second live session, 'this would indicate that the upset behaviour observed during the replay was not simply due to fussiness or natural decline in performance' (Nadel et al 1999, 166).

The results showed not only that the infants were engaged in the second live session, but that they were more engaged: 'A small standard deviation accounts for a homogeneously higher level of interest for the mother after replay than before' (Nadel et al 1999, 171). Nadel et al conclude that 'these results allow us to conclude that the changes in behaviour observed during the replay period were not a simple consequence of fussiness… This renewal of attention [in the second live session] can only be explained by her re-establishment of contingent social responses' (Nadel et al 1999, 172).

In case it might be thought that these experiments are limited to mother-infant, and thus not stranger-infant, interactions, subsequent studies have shown that the availability of such indeterminate interactions for the infant with his / her mother influence infant's responsiveness to (unfamiliar) others (see, e.g. Bigelow and Rochat 2006), i.e. there is a generalisable conclusion we can draw about the quality of interaction between persons (and not just between intimates), namely that this quality resides in large part in its indeterminacy – in not being a re-enactment of an allegedly necessarily shared script that determines what is appropriate and inappropriate to do.

Breakdown, then, against this account of interaction going well, is manifested precisely in the stiffening up, the fixating, the structuring of the interaction. To say this is not to say interaction cannot be so structured, and that this cannot be beneficial:

two parties in an interaction could treat certain behaviours as unchangeably required, monitoring their own and the other's behaviour in light of those requirements, and as part of the pursuit of some common aim. The point of the second-person perspective is to allow for this, but not allow it to dominate an understanding of interaction going well.

C. Autism and Prosopagnosia

There are some interesting echoes – which we cannot explore in detail here – between these studies on infants' sensitivity to contingency and the study of autism, as well as some mental disorders. In a paper to which we will return in the next Chapter, Søren Overgaard (2005) discusses some results of studies of autistic children conducted by Jonathon Cole in his *About Face* (1998). Cole reports that autistic subjects indicate that whereas they find static faces easy or at least easier to understand, they have great problems with ongoing interaction with another's face. As Cole says, 'faces make many subtle movements simultaneously, too many for autistic persons to process. A single expression, like a single musical theme, might be learned and understood by rote, but the complex, almost symphonic use of the face we all engage in was just too difficult' (Cole 1998, 107; quoted in Overgaard 2005, fn. 18). For the autistic person, 'the status face of a picture is controllable and able to be abandoned without consequence' (Cole 1998, 103; quoted in Overgaard 2005, fn. 18).

From the perspective of the studies on contingency, but also the above discussion on the quality of interaction generally residing precisely in its indeterminacy, these are interesting observations, for they locate the difficulties autistic children have in social relations with others in their need for something more determinate (their need to recognise some particular or specific pattern or concept, anger or joy as identified by some properties). Connected to this need for recognising an instantiation of a general model may be the lack of engagement with the other's actions *as* invitations – as precisely incomplete and thus opportunities to initiate or keep renewing an interaction.

Cole's observations above may also remind us – and thus offer another echo with the above findings of the importance of sensitivity to interactional contingency – of the patient described in Oliver Sacks's *The Man Who Mistook His Wife for a Hat* (1985), for in that book Sacks describes how his patient needed to interpret his face as if it expressed a pattern familiar to him, which he then associated with a concept

(anger or joy), and Sacks notes how difficult this made any interaction with the patient (Sacks suddenly became very conscious of his position as a doctor observing a patient, rather than interacting with another person; the case of 'Dr. P' is presented by Sacks as one of prosopagnosia, literally meaning difficulty in face-recognition).

Another example of the connection between autism and lack of sensitivity to contingency – and one that also illustrates the importance of a relation with difference rather than sameness – is the study made by György Gergely (2001). Gergely's paper responds to earlier studies (see Meltzoff 1990; and Meltzoff and Gopnik 1993) that claimed that infants, if given a choice between someone who is imitating them and someone who is not, will favour the person who is imitating them (thereby allegedly showing that infants are attracted to those who are 'just like me'). By contrast, Gergely's experiment – which involved showing the infant two screens, one of which involved perfect imitation and one which was imperfect (the first was generated by a computer, the second involved a human) – indicated that normal children (already by the age of 3 months) 'attended more to the imitation-based (highly but imperfectly) contingent image than to the perfectly contingent computer-generated one' (Gergely 2001, 417). Gergely colourfully refers to the second, imperfect imitation as 'almost like me' (and therefore preference for some difference rather than complete sameness). According to Gergely, this is related to autism because 'we hypothesise that in autistic individuals the normal shift at around 3 months (triggered by maturation or experience) in the target value of the contingency detection module does not take place (or not by enough). As a result, children with autism continue to invest in perfect contingencies throughout life' (Gergely 2001, 419). The effect of this, in turn, is that such children exhibit, for instance, 'intoleration to variation in routines'; an aversion and avoidance of social interaction due to the inability to attend to and process less than perfect facial and gestural dispositional cues; 'a lack of social responsivity'; and, in general, 'lack of social understanding' (Gergely 2001, 419-420). Gergely still speaks, in this context, of the dangers of these deficiencies for abilities to read minds, but from the second-person perspective what we can say is that the preference for 'just like me' as distinct from 'almost like me' shows a deficiency in being able to relate with someone different (at least to some degree) where the interaction is characterised, precisely, by a mix of the regular and the irregular.

If we return to Khushim, the hero of our epigraph, we can perhaps begin to see that Khushim seems incapable of precisely tolerating any 'variation in routine' – he

wants to follow the pattern given to him by his mother. He does not wait; he does not hesitate; there is no sense in which he issues an invitation to interact, or indeed looks for one (especially one he might not immediately recognise as an invitation). There is, in Khushim, no sensitivity to interactional contingency. The problem is not that Khushim does not have the ability to infer what others are doing from certain rules already known to him; the problem is that is much too willing to do that, and not able or willing to do anything else, e.g. precisely wait, be non-committally alert, and experience irregularity both before and in the course of interaction. More generally, then, one could say that autism is not the lack of an ability to learn and follow social norms; if the above is accurate, then it is the other way round – it is that autistic subjects, as a result of a lack in the sensitivity to interactional contingency, have to rely too much on social norms (on certain established, specific ways of relating to others) and, as a result, are unable to participate in the indeterminate spaces in which we interact with each other (from the second-person perspective).

Of course, to mention these possible connections here is not to neglect to acknowledge the complexity and difficulty in diagnoses of autism[222] and prosopagnosia, but just to point out that studying interaction from the second-person point of view might shed some light on the difficulties that certain persons (including those suffering from autism and prosopagnosia) might experience in relating with others in everyday interaction. Indeed, whereas, for those persons who participate in social interaction successfully, the mix of regularity and irregularity is welcome – it is precisely that mix that comprises the indeterminacy of the relation – for those with autism, such a mix is a distraction and an obstacle, with the ideal being a perfectly ordered, harmonious, synchronous and thus also predictable interaction. What is in one play and playfulness is, in the other, the order of a game.

---

[222] For more on a reading of autism in this spirit, see Hobson 1995.

Chapter 23. The Otherness of Others

As has been now noted several times, it is of great importance for the second-person perspective that the difference between persons in the interaction is maintained, and that no sameness is neither presumed as the default, nor projected as the aim. In this Chapter, the final one in this Part, we try to systematise the sense of 'difference' that the second-person perspective relies on, and also show how, so understood, it may affect certain currently popular ways of characterising 'social perception' or 'inter-personal perception'. We begin, first, with Overgaard's 2005 paper, mentioned briefly in the preceding Chapter. Having discussed Overgaard, we proceed to consider how Martin Buber's work may assist us in further understanding the kind of otherness of others we need to speak of in the second-person perspective. Finally, the Chapter contrasts Overgaard's and Buber's contributions with those of recent appeals to empathy and other perceptual or quasi-perceptual accounts of access to others.

A. Levinas and Wittgenstein on the Face of the Other

Overgaard's (2005) paper, which is entitled 'Rethinking Other Minds: Wittgenstein and Levinas on Expression', seeks to offer a balanced approach to the other minds problem, i.e. one that incorporates insights both from those who 'observe that very often we do have reliable knowledge about the feelings and thoughts of other persons' and those who 'say that the mental life of another is somehow inaccessible to us' (Overgaard 2005, 250). As the title suggests, Overgaard draws on Wittgenstein and Levinas to assist him in this task: Wittgenstein to help with the idea that we can have reliable knowledge of others, and Levinas with the idea that we cannot. Although the two may seem at odds at first, Overgaard argues that they can be reconciled. The way he does so instructive.

Overgaard does not attempt to reconcile Wittgenstein and Levinas by arguing that sometimes we do have reliable knowledge of others and sometimes we do not. Instead, he takes a much more interesting position:

> There is an important general particularity that characterises knowledge (or perception) of the feelings of others, in contrast to the knowledge (or perception) of facts about physical objects. Even when we do *know* what another is feeling, her feeling is presented to us *as hers*, not ours. This seems to make room for the idea that our cognitive grasp of her feeling, if one may put it this way, is less strong than the grasp we have of other types of fact, say

facts concerning the colours of material objects visible to us in broad daylight. Thus, even in the most favourable cases of 'direct experience' of another's feelings, we do experience something that we might, following Husserl, term the 'transcendence of the other.' In such cases, too, we experience that the other person eludes us in a particular way. This 'transcendence' is at the basis of the various kinds of doubt and uncertainty that we may have concerning the mental lives of others. Therefore, the most 'favourable' cases must already lay bare the foundations for both of the mention intuitions regarding other minds. (Overgaard 2005, 251-2)

Before we go on to discuss Overgaard's position in more detail let us recall that this thesis has addressed the other minds problem in a certain specific way (see Chapter 20). This thesis has rejected a reading of the problem (of what the problem demands) that makes it impossible to see how we might not need to 'know' other minds, indeed how a desire to know might be an obstacle to interacting with them (see also Chapter 22). As it is often set up, the other minds problem relies on examples where an interactional scene is described in such a way that agent A faces agent B, and agent B is described (by the theorist) to have a certain mental state (e.g. B is happy), the question being how agent A can have access to agent B's mental state (when, if ever, can agent A really know that agent B is happy). We have been pointing out that this way of setting up the problem is a misleading one, for it is much too static and third-personal, i.e. it neglects to make proper room for agent A's perspective as he or she interacts with agent B. When one does make room for that perspective, one sees that often the best agent A can do – and indeed, often what facilitates interaction in any event – is agent A orienting himself to what agent B might be feeling, but in a more flexible and tentative way, one that includes agent A offering his own intentions as invitations for interaction, calling for and waiting for agent B's reaction (which is experienced as incomplete, unfinished), and responding to that in turn, the point being that the response does not need to be based on any secure access to agent B's mental state (based, as it is often said to be as a result of this way of setting up the case, on an inference from some gesture or expression, e.g. a smile).

The way the problem of other minds is set up, as indicated in the structure of the above example, is that it makes seem as if there is a fact of the matter about what B feels, and that this is there waiting to be recognised by agent A – and this way of setting it up misses precisely what is tentative and interactive in agent's A relation with agent B. Agent A certainly needs to be active, but not active in the sense that he must treat B as an object to be 'figured out', but as another person to be related with,

interacted with, and this also means relating with agent B in a more tentative way (one not necessarily mediated by an interest in or need for knowledge or access). Crucially, this also means that agent A can appreciate agent B as perhaps being happy without this having to be mediated, personally or sub-personally, by, for instance, agent B smiling (or doing something else that agent A has previously treated as a rule for inferring the presence of happiness). The point, then, is that as a result of the interaction, agent A can come to appreciate a novel way in which agent B may appear happy. On the view we are resisting, this learning in the course of interaction is not possible (or is at the very least missed as a possibility). We shall be saying more about this when developing an account of cognition from the second-person perspective in the next Part of the thesis. But it is mentioned now, as it is important to keep this in mind when examining Overgaard's reply to the other minds problem.

Overgaard's starting point is to recall that Wittgenstein was, in many of his passages on sensations (e.g. on pain), battling the idea that sensations were something 'completely "inner" to the individual person, closed off from everything outer. This "self-containedness" view is part of the target of Wittgenstein's criticism' (Overgaard 2005, 254).[223] Putting things this way help us see that the following two views are compatible: first, my head hurting need not be by necessity tied each time to the same actions – e.g. rubbing my forehead, paleness; and second, my head hurting is not nevertheless somehow independent from my body. In other words, my body's gestures and expressions not being tied in some way to some state I may be in (some sensation I feel, e.g. pain) does not mean that my sensation is severable from, in the sense of being independent of, my body (precisely because I can experience a sensation in a novel way; my body, as it were, has more imagination than my mind). The importance of resisting the self-containedness view (the beetle-in-the-box view) is that it is only when we adhere to that view that we come to think that we need inferential access to another's mental state, i.e. we come to think that we must infer from observing some physical feature (e.g. paleness) to the presence of a head-ache (to the mental state of pain).

One can see how the commitment to inferential access can easily lead to scepticism (easily because all one needs is one case of a headache that is not

---

[223] The point can be put more positively, i.e. not in terms of anti self-containedness, but in terms of the other's expressions as incomplete invitations for interaction – of invitations precisely as communicative, meaning they cannot be already fully-formed objects passively awaiting recognition.

manifested by paleness, or one case of paleness that is not correlated with that kind of pain). In other words – at least in this context – scepticism does not need to be defeated, so much as acknowledged as showing us the limitations of any one way of linking up certain bodily expressions and gestures with certain sensations or states. Where scepticism goes too far is where it requires us to always be involved in the project of (implicitly or explicitly) justifying (even if only to ourselves) the way in which we see the other. The moment we accept this requirement, we forget that we can appreciate someone as perhaps being in a certain state, without this being mediated, personally or sub-personally, on some basis that can serve as a justification others might be likely to accept (because it is familiar to them, e.g. it is a rule of inference they are themselves committed to using). The point, in other words, is that most often we do not need more: when interacting with others, we do not need (indeed, if we did, we could hardly ever get around to interacting with others) to know for certain whether someone really is in such-and-such a state, and thus we do not to (though on occasion we can) claim or assert that someone is in such-and-such a state. The reason for this is that what matters (at least the great majority of the time) is not whether our own appreciations or claims / assertions are correct (whether we see what the other person allegedly really is feeling / thinking), but rather the quality of our interaction, e.g. do we respond to each other with sensitivity to interactional contingency, or do we get the job done (if we happen to be committed to performing some task together)? Part of this interaction can involve appreciating the other to be in some state, and this can occur without mediation (again, personally or sub-personally) by a rule of inference, and this is because the context of appreciation is independent from the context of justification.[224] But what is at stake in an interaction is not the correctness of our appreciation, but rather the quality of our interaction (where appropriateness can play its own important part in that, but does not exhaust it).

To repeat: the sceptic has a point; the sceptic reminds us that insofar as we rely only on some limited range of inference rules to make certain claims or assertions (about what another is feeling), this reliance may be misplaced. The important thing is to acknowledge this point, but to point out to the sceptic that we do not need to be ceaselessly making claims or assertions about what another is feeling / thinking, and that the sceptic thereby misses what is (mostly) at stake in an interaction, i.e. not the

---

[224] The way this is so is elaborated upon in Chapter 24.

correctness of our claims / assertions about the other (and thus, our 'knowledge of' or 'access to' the other), but the quality of our interaction with the other. The point, put differently, is that accepting the sceptic's terms, and going on to attempt to silence the sceptic forever, makes us miss the second-person perspective altogether.

Having resisted the self-containedness view, Overgaard's next move is to try to articulate the way in which we can think of the human body as 'expressive' of the mind (Overgaard 2005, 256). Overgaard understands that this cannot be too tight a link; in fact, although he speaks sometimes of 'typical' or 'characteristic expressions' (Overgaard 2005, 257), he also stresses – and here we already see a connection between Wittgenstein and Levinas – that: 1) 'expression is essentially non-static. A completely static face would not be an expressive face'; 2) partly as a result of (1), expression has 'limitless variability'; and 3) but that even (1) and (2) do not yet tell us how 'qualitatively different' expression is from the equally 'endlessly variable…movements of objects in space' (Overgaard 2005, 261). This is connected to how Overgaard reads Levinas because, according to Overgaard, Levinas helps in seeing the qualitative difference of human expression.

What Overgaard finds fascinating about Levinas is that unlike straightforward scepticism about the other – one that says that the other's mental state is inaccessible – Levinas focuses his remarks on the experience of the (expressiveness of the) face of the other. For instance, Levinas says: 'The face of the Other at each moment destroys and overflows the plastic image it leaves me, the idea existing to my own measure and to the measure of its *ideatum* – the adequate idea. It does not manifest itself by these qualities, but *kath' auto*. It expresses itself' (Levinas quoted in Overgaard 2005, 262). There are a number ways one could read this: one would be to point to the experience one can have of the inadequacy of any of one's ways of identifying or naming – as if freezing – what the other's face is expressing, for the other's face keeps on expressing; notice that to say this is akin to pointing to an experience of breaking an interaction, i.e. no longer seeing the other's expressions as invitations addressed to you, or as pleas for communication, but as performances to be read and described from a distance. Overgaard does not quite read it that way, though he also does not argue against such a reading. Instead, Overgaard points to the sense that we have, when interacting with the face of another (and not an object), that it is itself (and not me) that another's face expresses, 'That is, it unfolds in a certain dynamic that does not me at its source' (Overgaard 2005, 262). Our point is not to challenge

this, but to add to it: the way another's face unfolds in a certain dynamic that does not have me as its source is important, but it is equally important that the other's face has me as its addressee.

Let us say a little more about the way in which, for Overgaard, the sense in which the expression of the face is ongoing (dynamic) is connected to the sense of the other's face as the other's. Overgaard makes the following observations:

> It is crucial to realise that the mental life of another is not some sphere of 'inner' objects somehow displayed to my view – just as it cannot be reduced to external objects lying open to view. It is not something 'lying' or 'standing' at all; it is nothing 'present-at-hand', as Heidegger would say. On the contrary, the other's mental life is something that, unlike *any* object, presents *itself*, expresses *itself*, that is, an independent and inalienable source of meaning… On such an 'expressive' model, in contrast to the 'beetle model', the inaccessibility or 'transcendence' of the other person is not due to the fact that she has perceptual access to private entities that I cannot perceive. Rather, her inaccessibility has to do with the circumstance that she is the one expressing *herself* in her face, thereby always retaining the ability to overturn every firm conviction I come to form about her. (Overgaard 2005, 262-3; original emphasis)

There are a couple of things to note here: first, when I experience another's expressive face, I can experience it richly (e.g. as appearing sad or angry to me) as well as experiencing it as the face of an other (the way it appears sad or angry to me need not be familiar to me as a way I would express sadness or anger; it is very much the other's expression I can experience); and second, I can have doubts as to whether the manner in which I experience the other's face is how the other might explain it to me, but these doubts are better understood as a reminder of the difference between me and the other, reminding me that the other's feelings have a source other than me (the point is that my doubts are not about which general scheme or facial pattern to apply to make sense of the other, but are rather sourced in this being an other, someone unlike me, who is appearing to me). However, just because I can have such doubts, especially if I perceive the other's expression to go in a direction I would not have anticipated,[225] does not mean that we should introduce a gap between the body and the mind, and one mind and another – where such gaps would mean that the body is a puppet of the mind, such that the intentions / mental states remain inner, thereby

---

[225] We probably have all had the experience of thinking at first that someone is crying, but going on to think they were laughing – hence, perhaps, the expression 'crying with laughter'.

requiring one person observing another to infer the presence of some inner mental states on the grounds of the outer.

Overgaard's 'reconciliation' between Wittgenstein and Levinas is an instructive one, and we can build on his sensitivity to the rich and dynamic expressiveness of the faces of others. Indeed, we need to build on his account for in itself it is not interactive enough. We see that it is not interactive enough when we see that, for instance, Overgaard retains the opposition between the first-person and the third-person (see Overgaard 2005, 263-4).[226] He does this for a good reason, i.e. he wants to show us (following Wittgenstein and Levinas) that seeing the difference between the first-person and the third-person means that we can appreciate the difference between something appearing to me as my own (my feeling) and something appear to me as 'her feeling' (Overgaard 2005, 264). As he points out, this does not mean we need to assign any special privilege to 'first-person access'; instead, what it means is that it helps us fasten onto the different experience we have when we experience others' feelings (as opposed to our own; Overgaard 2005, 264).

Overgaard's observation about this difference is an important one, but because he does not make room for a distinctively I-you, as distinct from an I-her, encounter, he also retains some of limitations of an opposition between the first- and third-person. As we have been pointing out, in an interaction understood from the second-person perspective one experiences the other in an indeterminate space where one interacts with them, and where, as a result of being in such a space, one both invites the other to interact, waits (and thus suspends one's ability to claim or assert that another is feeling or expressing such-and-such a thing), and experiences the other's actions as incomplete and unfinished opportunities and renewals of interaction (the unpredictability of the direction in which the interaction could go is connected to one's susceptibility to being moved by the other in a way one cannot control or imagine in advance).

Overgaard's observation about the way in which we can experience another's expressiveness and feel it to be hers (the other's) captures some of the tentativeness and uncertainty, some of the interactive tension, in my experience with the other (in this sense, it helps us capture that otherwise mysterious-sounding notion of the

---

[226] In reverting to talk of 'the first-person' and 'the third-person' here we do not mean to endorse this way of using these terms. Again, the notion of the second-person perspective as used in this thesis is distinguishable from the way that Overgaard uses these terms.

'otherness of the other'), but it needs to be put in the context of interaction understood from the second-person perspective. When it is not, it risks defining the distinctive perspective of second-person out of existence, e.g. as when it is read to say that, in order to interact with an other, I can and need to have 'the firm conviction' (to follow the quote above) that someone is feeling such-and-such, albeit with the provisos that I experience that conviction as targeting someone other than me and, as a result of that, remain open to changing my view. The problem here is that even with the proviso that the relevant feeling is the other's (and not mine), the account reduces what could be an interaction to a context where I make a claim or assertion 'about' you (even if this is a claim or assertion I am prepared to change as a result of what you tell / show me).

To make this point is also to remind ourselves of how our inheritance of a certain philosophical problem (the problem of other minds in this case), read in a certain way, can make our own philosophy miss something important, for the other minds problem can all too easily be read as a problem concerning the status of my claims or assertions, making it seem as if when I interact with others I must always make claims or assertions, which I can back up by appeal to some grounds. The second-person perspective allows us to see the limitations of this approach. Put differently, the second-person perspective allows us to avoid shoehorning our account of how the mind learns to interact and does interact with others into the rather static mould of the status of my claims or assertions about an other's mental states. When I interact with an other, I do not (always, by necessity) need to make claims or assertions about the other (to 'know' or have 'access to' the other) – indeed, quite often, in order to interact, I cannot make claims or assertions about the other, for as soon as I do so, I step outside of interaction understood from the second-person perspective. Certainly, there may be moments in any interaction where I do make claims or assertions about the other, but what is important is that we do not understand interaction as necessarily requiring, or being exhaustively dominated by, such a form of relating to the other.

B. I and Thou

Overgaard's sense of the otherness of the other (of the face both revealing and concealing itself – expressing itself, but also expressing itself as its own, and thus unlike mine) can be further supplemented by other ways in which theorists have

spoken of my encounter with the otherness of others (and thus help us understand the specific sense of 'difference' of importance to the second-person perspective).

One of the traps to avoid here is thinking that the other is different, and only ever different, in some specific way, some way that we can fit into some scheme, and classify as an instance of that particular difference between persons (of course, we can have this experience, but the point is not to reduce the otherness of the other to that). One figure who can help us in this respect is Martin Buber, whose *I and Thou* (1970) is, though difficult, a rich source of insight about the experience of otherness. Unlike Overgaard, Buber retains a difference – and it is a vital one for him – between the second- and third-person. What Buber wants to show is the distinctiveness of the I-You encounter from an I-It (including an I-She/He) encounter.

A good way of understanding his approach (as set out helpfully in Arvidson 2003) is to see in what circumstances Buber would argue that an I-You encounter has been reduced to an I-It encounter. For instance, Buber writes: 'I can abstract from him the colour of his hair or the colour of his speech or the colour of his graciousness…but immediately he is no longer You' (Buber 1970, 59). Arvidson calls this 'singling-out': pointing to / identifying / extracting or abstracting a particular feature (see Arvidson 2003, 86). Another example is what Arvidson calls 'synthesising', as when I see you as part of some group, or 'constituent of a theme', e.g. 'I can envision You now as a member of a family or as part of what is exciting in the room. To that extent I no longer encounter You' (see Arvidson 2003, 86-7). In this respect, Buber says: 'I can place him there ['in any Somewhere and Sometime']…but immediately he becomes a He or a She, an It, and no longer remains my You' (Buber 1970, 59; quoted in Arvidson 2003, 87).

There are two things we can observe about this: first, that asserting the 'particularity' of the other can be misleading, for if we think of particularity as being specific in some way, then we in fact define out Buber's sense of 'You', which is not particular so much as indeterminate; and second, we also see that the issue is not one of 'directness' or 'immediacy', i.e. there is no idealisation here of an encounter that is easily and perfectly meaningful and harmonious, but rather something like an experience of you as your own source of action and meaning (not one that I control or impose on you, but one to which you can invite me).

It is important to stress this second point, for otherwise we might be charged here with endorsing that which we (in Chapters 20-22) distanced ourselves from,

namely that sense of the ideal connection, a kind of perfect matching, of something pure and un-mediated. In fact, what we have been advocating for is (almost) the opposite: a mixture of engagement and disengagement as part of a relational (interactional) dynamic, and thus not something experienced immediately as meaningful and fulfilling, but rather as an indeterminate process in progress – something to which both parties are committed, something they care about, but not something without (indeed, precisely with) tension and uncertainty. What Buber allows us to see is that in respecting and maintaining difference when describing interaction from the second-person perspective, we must avoid understanding that difference as necessarily different (and thus intelligibly different) in some specific way (whether that be as a singular feature, or as a feature in some general scheme).

One of the principal points, then, to make about the otherness of the other from the second-person perspective on interaction is that it is not immediately or easily intelligible, but (almost) the contrary: that the other need not be so intelligible to us, partly because in interacting with the other we do not need to rely on such specifiable intelligibility, but also partly because an element of our experience of the other is the sense in which the way the other 'is' is indeterminate. Thus, it is not that the other is immediately intelligible to us in a certain way (say out of a stock of available, familiar categories we can impose on her), or that the other is indeterminate in the sense in which she can be understood through a multitude of (available, familiar) categories (that somehow may apply to her, depending on the circumstances), but rather that the other – in the course of our experience from the second-person perspective – remains a centre of autonomy, and thus not also someone to be understood, figured out, an object to make claims or assertions 'about', but rather another person to interact with, to be surprised by, be moved by, and also to surprise and to move in turn – a process that, precisely when it is going well, involves a lot of stepping on toes, fits and starts, (some) invitations not taken up, (some) opportunities missed.

C. A Note on Immediacy

It may be helpful to pause here for a moment to consider talk of 'immediacy.' Reference to immediacy tends to come in two forms: first, as an attempt to avoid the necessity for postulating an inferential process (at the phenomenological level) in our sensory encounter with the world; and second, as an attempt to convey the kind of

experience we can have of seeing something without that seeing being accompanied by doubt (the two are related, as we shall see in a moment). From the second-person perspective the idea that we can see the world and not doubt 'what' we are seeing is problematic; the point is that when we interact with another, we need not in any way see / feel that there is a 'that' to our experience (e.g. that the other is sad or happy) – put differently, we tend to characterise our experience as having an object (a 'that') when we experience some reason to doubt, some reason to acquire evidence more systematically, to prepare ourselves for a justification (acquire a 'firm conviction'). We need not experience this kind of doubt when interacting with another (indeed, it is often the case that we cannot experience this kind of doubt – a kind of self-reflective doubt, a doubt based on whether we are correct in what we see), but this does not mean that we do not experience any doubt at all when interacting with others. Perhaps the key thing to see is that the doubt must be an *interactive* one (not a self-reflective one, not one concerned with the status of our own experience). Indeed, as we have been describing interaction from the second-person perspective, there needs to be some experience of hesitation, tentativeness and even uncertainty, but this should be understood as part of a flexible orientation to the other (including an experience of the otherness of the other – of a difference that cannot be assumed as making itself intelligible to us immediately in accordance with what we already find familiar). This flexible orientation to the other requires a kind of sensitivity to contingency (to how the other might be), but this sensitivity is not structured by the kind of doubt suggested by talk of immediacy: we do not, as it were, by default (and thus perhaps in some sense sub-personally) see / feel (unless we are given reason to doubt) that the other is doing / feeling such-and-such a thing.

One can see how the two forms of immediacy are related here: in order to fend off the possibility of doubt, experience is described as having a certain structure (with already in built self-justifying mechanism that operates automatically – automatically prepares a ground that can be appealed to if questioned). The point being made here is that when we interact with an other, it is not this kind of doubt that accompanies our interaction with that other; of course, we can stand back and reflect upon how and whether we should judge the other as being (in some specific way), but this is an exercise that is no longer (a central, dominating) part of a second-person interaction.

We must, then, be careful of talk of immediacy, for it can make us describe experiencing the other in interaction (understood from the second-person perspective)

in a way that, in fact, loses sight of what the second-person perspective reveals as important to interaction. It is precisely the advantage of the second-person perspective that it allows us to characterise the experience of other persons in social life in a way that a social agent cannot quite be fully prepared for. We might say that what emerges as important from interaction understood from the second-person perspective is not doubt about a person's own judgements, but a kind of interactive doubt, i.e. the sense of uncertainty, hesitation and tentativeness persons can experience when relating with another person. When we see this we also allow ourselves to see that persons need not be pictured as having learnt, and internalised, and now be applying general categories to make others intelligible in some specific way, which in turn allows us to see how the way persons learn and interact with others is inherently more flexible than we otherwise, under the pressure of self-reflective doubt, characterise it as.

With this sense of the otherness of others in mind – and thus as an experience situated from a second-person perspective of interaction – let us briefly consider how what we have said here is somewhat at odds with some of the ways in which persons' understanding and perception of other persons has been accounted for.

D. Empathy

The most recent, full-scale defence of empathy as 'the epistemically central, default method for understanding other agents within the folk-psychological framework' can be found in Karsten Stueber's recent book, *Rediscovering Empathy* (2006; the quote is at 6). Notice the reference to 'epistemic'; this is very important for Stueber, and he understands this as requiring any account of our 'understanding other agents' to show that we can have 'justified true beliefs about the mental states of other people' (Stueber 2006, 17). Empathy, then, is presented as the solution to the other minds problem, which is understood as requiring a 'method [that] will provide us with genuine knowledge of the other person's mental states' (Stueber 2006, 18). We have been arguing that situating persons in relations with others in this context of justification has the effect of skewing, and ultimately narrowing, our description of how we learn to interact and how we do interact with others (understood from the second-person perspective). We see this narrowing also in Stueber's account of empathy, which offers two forms of empathy – basic and re-enactive empathy – both of which are too static and rigid as sources of the ability to relate to others (again, from the second-person perspective of interaction).

Basic empathy refers to 'mechanisms that underlie our theoretically unmediated quasi-perceptual ability to recognise other creatures directly as minded creatures and to recognise them implicitly as creatures that are fundamentally like us' (Stueber 2006, 20). There are two things to note about this: first, the reference to 'directness' (by theoretically unmediated Stueber means to say non-inferential – not requiring a theory to be applied); second, the reference to the need to see others as 'fundamentally like us'. Thus, basic empathy fails both to make room for the kind of interactive doubt we referred to above, as well as failing to see how we can (indeed need to, for second-person interaction to go well) experience the otherness of others in the course of interacting with them (and thus to see them as fundamentally not like us). The epistemic and non-interactive spirit of basic empathy is also visible in the examples of it offered by Steuber: 'Basic empathy…allows us to recognise, for example, *that* another person is angry, or *that* he intends to grasp a cup' (Stueber 2006, 21; original emphasis). We have been arguing, to the contrary, that in the course of second-person interaction we need not, and to some extent cannot (in order to facilitate interaction), be making claims or assertions 'about' the other person in this rigid, static, stop-start manner.

Re-enactive empathy, in turn, is a form of empathy that enables us 'to understand why that person is angry, or why he responded to a particular situation in a certain manner… [it enables a] complex understanding of another person's behaviour' (Stueber 2006, 21). The way it works, according to Steuber, consists in 'using our cognitive and deliberative capacities in order re-enact or imitate in our mind the thought processes of the other person' (Stueber 2006, 21). It is this second form of empathy that is properly said to belong to the 'folk-psychological context' in which, says Stueber, 'we view each other as rational agents', i.e. as agents who act for reasons (Stueber 2006, 22). These reasons are susceptible to folk-psychological explanations (indeed, they count as reasons only because they are so susceptible) and they also form the data according to which we can evaluate whether they are rational – for Stueber, our evaluation of them as rational proceeds in accordance with the norms of rationality, though with the proviso that because of the contextualised nature of our judgement (as to whether someone is rational) we need not think that

someone's violation of the norms of rationality automatically means they are irrational (Stueber 2006, 23).[227]

Notice that in both basic and re-enactive empathy, one person faces another as an object to be understood: first, in the alleged basic sense as to 'what' they are doing or feeling, and second in the allegedly more complex sense of why they are doing what they are doing and whether doing that is a rational thing to do. There is no sense in either of these two processes of anything interactive – no sense, for instance, of the way in which the other's actions may move me, entice me, call me to respond, invite me to a relation, and be, as a result of that, incomplete, inherently indeterminate, more like opportunities for response than objects to be defined and evaluated.

Stueber does argue that his conception (of our use of folk-psychological explanations) is an 'engaged' one, and he argues specifically against a 'detached' conception of that use (in chapter 1 of the book). But what Stueber means by 'an engaged conception of folk psychology' is that 'folk-psychology [is]… a theory that is adopted from the point of view of agents who in experiencing their agenthood have to view themselves in folk-psychological terms and who use the same terminology to explain and predict the behaviour of other human beings who they conceive of as agents like themselves' (Stueber 2006, 46). In other words, 'engaged' means treating the other as alike in the sense that the same schemes of intelligibility and the same standards apply to oneself as they do to the other. Being engaged for Stueber has nothing to do with a robust sense of interacting with another, which, as has been argued in this Chapter, in fact demands relating to, including respecting and maintaining, the otherness of the other.

Stueber could here reply: but surely we can stand and define, judge as intelligible, and judge as rational (or not), some other person's action. Certainly, this is true, but the point is that if we emphasise this context in which we may face others, and fail to make room for seeing interaction from the second-person perspective, we risk describing how we learn to relate, and how we do relate to others, in a very

---

[227] There is a tension here between understanding others as intelligible in line with folk-psychological explanations of them and evaluations of them as rational partly in accordance with rational norms, for would not our judgement of intelligibility render the judgement as to rationality superfluous? Is that not the point of the application of folk-psychological explanations? Who, and why, would need to go further and appeal to rational norms? Stueber seems to want the application of folk-psychology to be merely an intelligible stage, one that makes it susceptible to further evaluation as being rational or not – but is not the point of the application of folk-psychology that it sufficiently rationalises the behaviour for us (that it makes it precisely unnecessary to proceed to a third stage)? We shall leave discussion of this issue in Stueber's work for another occasion.

narrow, i.e. static and rigid, way, one that, in the end, that makes it impossible to account for how we could interact with others.

Recall that Stueber's definition of basic empathy refers to it as a quasi-perceptual faculty, and indeed, recent defences of the role of empathy have characterised it precisely as consisting of a 'perceptual theory of other minds' (Zahavi 2010, 292, in this context speaking of Max Scheler). Stueber, as we have seen, goes further, in that he limits empathy so understood to something 'basic' and calls upon another kind of empathy – re-enactive empathy – to point to what enables an allegedly more 'complex' understanding of others. He shares this move with others, e.g. Alfred Schutz argued (as summarised by Dan Zahavi) that although it is 'permissible to say that certain aspects of the other's consciousness, such as his joy, sorrow, pain, shame, pleading, love, rage and threats, are given to us directly and non-inferentially, he denie[d] that it should follow…that we also have a direct access to the why of such feelings', adding that 'in order to uncover these aspects…we also have to rely on interpretation, we also have to draw on a highly structured context of meaning' (Zahavi 2010, 297; see Schutz 1967).[228]

Zahavi has Schutz referring to direct access to a wide range of emotions (or mental states of the other), but Stueber restricts his notion of basic empathy to a narrower range of emotions (again, basic emotions like fear). Stueber wants his notion of basic empathy to give us direct access to basic emotions 'based merely on a perception of their facial expressions' (Stueber 2006, 140). We will return to this in a moment, but notice that in Stueber's examples of basic empathy, he does not restrict them merely to recognising certain basic emotions, but also intentions: '*that* he intends to grasp a cup' (Stueber 2006, 21; original emphasis). Interestingly, Schutz also allows for this – i.e. for perceptual, immediate / direct, grasp of another's goals in his actions – but only if the two persons 'share[] a motivational context [sometimes also referred to as shared situations or pragmatic contexts] where our respect streams of consciousness are interlocked, immediately affecting each other' (Zahavi 2010, 297; see also 302, and see Schutz 1967, 170).[229]

For Steuber, the way basic empathy works – both in the sense of recognition of basic emotions as well as recognition of intentions – is that we have 'cross-modal

---

[228] It has to be said that it is surprising, given some of the similarities between his view and Schutz's, that Stueber does not even once refer to Schutz.

[229] This requirement by Schutz of a 'shared motivational context' is also what restricts the use we can make of his work for the purposes of articulating a second-person perspective on social life.

mechanisms underlying our imitative capacities [which] allow us to encode other agents as being "like me"' (Stueber 2006, 143). According to Stueber, 'the infant is wired by nature to encounter other human beings as belonging to the same set of "objects" that it belongs to. It sees other human beings as similar to itself and understands itself to be more similar to them than to any other "objects" it encounters or interacts with in the world' (Stueber 2006, 144-5). Notice that we can say that we encounter persons differently to objects without needing to say that I must see them as like me – in this respect, the notion of similarity is ambiguous here: it works both to distinguish the other as a person and not an object, and to establish someone as like me, i.e. someone who has the same emotions as I have recognised myself to have. Notice, too, that for Stueber what infants are learning are 'mental state concepts' (the concept of fear, for example) which they then apply in a particular instance to another person: 'I understand mental state concepts as used by another person in light of perceptual similarity spaces due to the mirror neuron systems as being applicable to both myself and other persons, particularly those that are closely tied to certain behavioural manifestations such as is the case with certain emotions' (Stueber 2006, 145).

There are a number of observations one can make about this analysis: first, it is not clear, as we have seen above (in Chapter 20), that mirror neurons are about imitating (in this sense of passively mimicking) another, but may involve attempts to establish an interaction (attempting to repeat the other's actions is a way of getting their attention, and inviting them to relate); second, it is not clear how our capacity to imitate (whether imitate in an interactional sense, or mimic) allows us to learn 'mental state concepts' – we see the difficulty Stueber here faces when the only example he provides (of learning, rather than recognising) is one where the infant hears the name of a concept being associated with some expression of another: 'I grasp implicitly that, when my mother says about another person that is angry, she is using a predicate that also has conditions of application in my own case' (Stueber 2006, 144) – this greatly restricts how infants learn (and is furthermore in contradiction with Stueber's wish to see basic empathy as pre-linguistic); third, it is not clear what the relationship between 'mental state concepts' and 'similarity spaces' is, i.e. whether the mental state concept in question is one identified by certain specific facial patterns (such that it could not be appreciated in any other way), or whether there is a loose grouping of facial patterns (when Stueber uses 'similarity', he seems to suggest that I see someone

else as similar to me, and this then enables me to apply the same specific patterns I associate with myself to the other; this suggests a tight rather than loose grouping between patterns and concepts); fourth, all this makes it difficult to see how persons can learn from others (how the experience of another can help someone discover a kind of intention a person can have – for it restricts all understandings of others as a matter of projection from oneself to another); and fifth, it remains static and non-interactional, in that the person (including the infant) is seen as imposing certain categories onto another who exists passively in a frame, as if waiting to be defined and classified (one can see here again how this loses sight of the relational sense of the otherness of the other experienced in interaction from the second-person perspective).

E. Perceiving Others

We mentioned above that Stueber goes beyond the 'basic' 'quasi-perceptual' reading of empathy, and relies on re-enactive empathy for something allegedly more complex. For those who emphasise the perceptual credentials of empathy, the perceptual mode is very complex, or as complex as is needed, including complex enough to give perceptual access to reasons for action.[230] Certainly, when it comes to perceiving others' emotions, philosophers have not restricted themselves to basic emotions (even if the headline examples often concern perceiving another person's anger: see, e.g. Dretske 1973; and more recently Cassam 2007, ch. 5; Stout 2010; and McNeill 2010), referring to 'misery' and 'self-pity' (see Smith 2010, 741), 'pain' (McDowell 1978; see also Seemann 2008),[231] 'sorrow', as well as (and here perception is used broadly in the sense of being multi-modal) 'the anxiety on someone's face…the trepidation in her handshake…the exuberance in her voice' (Green 2010, 45, and originally in Green 2007). In this debate, the basic issue is whether the properties that identify some specific emotion (or mental state) present themselves directly to our senses where there is no need for an inference, or whether the basic pattern is what is directly perceived and then an inference is made to what the pattern means (i.e. that the person must be, e.g. angry).

---

[230] For a recent account of perceiving reasons see Church 2010.

[231] Seemann does say that he wants his account to be 'interactive', in a way that maintains the difference between self and other (Seemann 2008, 249), but he still seems to rely on situating interactions in shared social environments (e.g. Seemann 2008, 258); see also Seemann 2010, where he also speaks of 'sharing feelings' in the context of paying attention to 'the same features of the perceptually presented environment' (Seemann 2010, 169).

From the perspective of this Chapter, the problem with the way the debate is set up is two-fold: first, it ties specific emotions too tightly with properties that allegedly identify them, whereas it is better to say that the reference to properties is something we do in the context of justification and systemisation, and not in the context of appreciation (more on this in Chapter 24); and second, it is too static, and not interactive enough, in that it presents the other as an object of my recognition rather than my partner in interaction. Put another way, the problem with the debate is that its focus is on precisely the status of my access to or knowledge of the other, where the focus of the (positive) efforts in the debate are on showing how it might be possible for a person to have access to or knowledge of the other without engaging in an inferential process.

As we have arguing, though, the very problematic here is one that makes us characterise experience as implicitly concerned with its own status, and thus makes us think of experience as a series of claims or assertions about how things are. What we have been saying is that this approach is too static and too anxious (in a self-reflective manner), and one that makes us neglect the kind of picture of cognition that opens up when we take the second-person perspective on interaction seriously. In this respect, Stout's recent contribution to the debate – Stout 2010 – is interesting, for he wants to take interaction seriously (referring, for instance, to Reddy 2008), and perhaps as a result speaks of 'processes' rather than static recognition of states; however, as his paper raises issues better dealt with in the next Part, we shall postpone a discussion of his contribution till then (see Chapter 25).

What we have said here about the debate over perceptual access to others applies also – though it is set in a different problematic – in the context of social psychological theories of social perception, i.e. these tend to be about applying general schemes (whether stereotypes or roles, or various other kinds of schemes of intelligibility) onto particular others.[232] We will be examining why the appeal to generality (in the form of the application of general models to particular others) is not necessary in the next Part, but here the point to make is, once again, that the relation is here described as too static: as one involving a one-way imposition of a category onto another.

---

[232] See, e.g. Kinda 1999.

One recent exception, which may offer helpful resources in further developing a distinctively second-personal approach to social cognition, is Frédérique de Vignemont's paper on 'Frames of Reference in Social Cognition' (2008). De Vignemont's contribution positions itself in the mind-reading tradition (thereby still assuming we need to mind-read), but otherwise is very alert to the differences between what de Vignemont refers to as 'social interaction' on the one hand, and 'social observation' on the other.

De Vignemont defines social interaction as 'oriented toward persons that are considered as a "thou", as another agent that cannot be treated as an object', where the interaction 'can consist of affecting the other's behaviour and experiences', such that 'we can adjust to each other, and I can test your reactions, to see how your assumed motives produce what you say' (de Vignemont 2008, 91). There are some aspects of the approach that are not necessarily congenial to how we have been articulating the second-person approach, including reference to 'immediacy' in the face-to-face encounter, and the idea that 'the environment can be guaranteed as a common one, shared in experience' (de Vignemont 2008, 91), but these ideas may be divorceable from the thrust of de Vignemont's social interaction picture.

The applicability of de Vignemont's view will, in the end, depend on the details of description her examples, which in the above-mentioned paper are sparse: for example, although de Vignemont describes a case of an I-You and not a I-He/She encounter (in the latter, an agent sees a dispute between Peter and Mary, and in the former, the agent is addressed directly by Peter), what she stresses about this is what she calls 'the egocentric frame of reference', namely that 'to react properly to Peter, I need to know that it is with me and not anybody else that he is angry' (de Vignemont 2008, 95). This is a valuable insight – it helps us get at the experience of being addressed – but it does not quite get at the question of how I appreciate what Peter might be doing as an incomplete and unfinished invitation / opportunity / call for me to respond. It is, instead, as if I perceive myself as an object of anger from Peter's perspective; the point being that the 'object' could just as well be, say, a vase about to fall on me – there is no interaction, or it is very minimal.

In this last Chapter especially, but throughout this Part, we have already found ourselves referring to the kind of cognition persons need to interact from the second-person perspective. Certainly, our focus has been on understanding interaction itself, including when it goes well and when it breaks down – from the second-person

perspective – but such an understanding is not completely divorceable from elements that are perhaps more accurately described as features of how persons experience each other in interaction so conceived (again, this is particularly clear in the element discussed in this Chapter, namely the otherness of others). In the next Part, we will turn in detail to the kind of cognition needed for interaction from the second-person perspective.

To recap: we have seen that the second-person perspective on social life, including its normative dimension, is one that sees it as a mix of regularity and irregularity (though where, for the sake of making the argument more striking we have emphasised irregularity), where an interaction so conceived breaking down becomes a matter of it becoming too regular, too ordered, or too patterned. Persons without the ability to participate in social life so conceived – such as, arguably, persons with autism – in fact aim for and are attracted to that which is perfectly regular, ordered and patterned, and therefore also predictable. These persons are attracted to sameness rather than difference. When social life is explained, there is a risk we will adopt a kind of autistic perspective, i.e. assume or posit an order, whether purely regular or regular plus normative attitudes, or regularised rules, etc, and explain persons' abilities to participate in social life as a matter of conforming to that order (e.g. by coming to know and internalising the rules). Taking the second-person perspective, however, allows us, as theorists, to avoid that result: it allows us to make room for both regularity and irregularity in social life. It thereby also allows us to make room for a wide variety of experiences of that irregularity that persons have when interacting with each other (even in quite specialised contexts, such as professional diplomacy). The point here is not to introduce chaos where there was order; it is to get over that dichotomy, and to see that social life, including in its normative dimension, requires us to see, and recognise the importance of, that mix or balance of fluency and hesitation, ease and difficulty, confidence and doubt, clarity and ambiguity, and assertiveness and tentativeness. We now need to zoom in on the cognition that enables persons to participate in social life understood from the second-person perspective.

Part V. Cognition from the Second-Person Perspective

We have seen, in Part III above, how the governance view's insistence on the necessity of determinants of appropriateness and inappropriateness has certain implications for an understanding of how the mind learns and works. In this Part, we wish to show how the mind learns and works from the second-person perspective on social life. It is also in this Part that we see how the second-person's perspective on social life sets up the possibility of an arguably more nuanced, more careful, and perhaps ultimately more accurate description of the cognitive abilities persons need to participate successfully in social life. In other words, seeing what we see as important for social life through the second-person perspective, we begin to see a need for a fuller array of cognitive abilities – and, especially, not just those abilities that are describable as having developed and operating on the basis of acquired and internalised normative resources.

Cognition from the second-person perspective is a partnership between 'appreciation' on the one hand and 'attentive silence' on the other. These are technical terms used in specific ways in this thesis, and it is the task of this Part to systematise the glimpses offered of these terms in the Parts above, and to give an account of how they work. In doing so, this Part draws on literature and examples that are not necessarily focused on social life – some of the examples, for instance, refer to person-to-environment relations, rather than to person-to-person relations. This is intentional, for it is claimed that the cognitive abilities referred to here – again, the partnership between appreciation and attentive silence – are, although particularly important for person-to-person relations (from the second-person perspective), also possibly applicable to person-to-environment relations. In other words, although there are many specific features of person-to-person relations – as the second-person perspective has also sought to show in Part V above – some of the cognitive abilities needed in those relations are transferable, honed and developed in part in person-to-environment relations.

In case too much is expected, it has to be pointed out that the discussion below is limited. It does not seek to present all the cognitive abilities persons might need for participation in social life, nor even all the cognitive abilities required for participation in social life from the second-person perspective. For instance, nothing is said of other sources of constraints on action or interaction, such as the constraints

of certain institutions (e.g. asylums), or features of the physical environment (as emphasised in externalist accounts of the mind, e.g. in the work of Andy Clark). The point is not to be comprehensive: it is, instead, to offer a sense of how much wider we need to cast the net when it comes to abilities that persons need to participate in social life, at least when we do not exclude, from our understanding of the social, the second-person perspective.

Given space restrictions, little is also said here concerning where the two abilities – appreciation and attentive silence – fall in the tradition of the philosophy of mind. Having said that, it may be best to see these two cognitive abilities as falling somewhere between a continuum, with belief at one extreme and perception at the other. At the risk of over-generalisation and over-simplification, it might be arguable that philosophy of mind has been caught in a dichotomy between belief (thought to be truth-seeking, possibly necessarily conceptual and therefore intelligent) and perception (thought to be, by contrast, sensory only, and thus naïve and blind). The description of appreciation and attentive silence, then, might be characterised as a plea for a wider array of abilities, more finely distinguished, and falling in between belief and perception. Of course, the above statement is made very tentatively, in acknowledgement that such claims, to be taken seriously, would need to engage properly with the philosophical tradition – a task regrettably outside the scope of this thesis.

There are two Chapters in this Part. The first, Chapter 24, turns to appreciation, distinguishing the context of appreciation from the previously-referred to contexts of justification and systematisation, and further developing the notion of appreciation by reference to the work of Frank Sibley in aesthetics. This Chapter also applies what is said about appreciation generally to normative appreciation. Briefly, when we appreciate, including appreciate normatively, we do not make claims or assertions as to how things are; instead, we tentatively appreciate something or someone as perhaps being so-and-so (including perhaps doing something or being appropriate or inappropriate), without the need for prior knowledge of properties that might form a ground or basis for a justificatory or systematisable claim about that something or someone being so. The point – which is important if we are to see interaction, from the second-person perspective, as a mix or balance of the regular and the irregular – is that we can relate to others (and possibly the environment) with minimal assertiveness, with a sense of non-committal alertness, and thus in a way that

we ourselves would not have anticipated nor been prepared for. This is important for participating in social life from the second-person perspective, for, as we have analysed it in Part V above, persons participating in social life need to be able to be surprised, and respond in ways new to them, to the otherness of others (to a difference that is not reducible to something familiarly different). The cognitive ability of appreciation, then, arguably captures some of the flexibility and tentativeness required to participate in interaction with others that is always, to some extent, unpredictable and ambiguous.

The second Chapter, Chapter 25, describes what is meant by 'attentive silence'. In brief, attentive silence refers to a mode of relating to the environment or other persons in which we do not either make (justifiable or systematisable) claims or assertions as to how things are or appreciate that things might be so. In short, attentive silence is closer to the perception extreme (on this imaginary belief-perception continuum) than appreciation is. Although appreciation is world-involving – it is a form of response to what is engaged with that is outside the person – attentive silence is even more so, and that is because it is even less assertive and less committal than appreciation. Attentive silence, if you like, is a form of taking in, of breathing in, the other (or the environment), without making either a justifiable or systematisable claim, or even an appreciation about how that which is taken in might perhaps be appreciated to be. It is, however, not quite naïve or blind perception – it involves a rich immersion in others or the environment as it is experienced, and one that could be reported on later if the person was asked.

What is important to see, and stress from the outset, is that the view of cognitive abilities developed here is one that depends on both working together: the mind learns and works in a way that involves the partnership of appreciation and attentive silence. These two cognitive abilities are different to belief and perception ultimately because they are less assertive and less committal than belief and perception are typically described as being. That is not their weakness – it is their strength. It is the claim of this thesis that it is those cognitive abilities working together in partnership that persons need in social life as understood from the second-person perspective. That, to repeat, does not mean that persons do not need other abilities – certainly they do – but it does mean that we must make room for these abilities as well. As has been argued above, as a result of how the governance view

approaches the normative dimension of social life, it excludes the possibility of seeing and describing the two cognitive abilities described in the two Chapters below.

Chapter 24. The Context of Appreciation

Throughout this thesis, reference has often been made to three 'contexts': justification, systematisation and appreciation. All three may be thought of as ways of experiencing the environment and others, but there is a distinction of vital importance between the first two contexts and the third. It is the task of this Chapter to articulate this difference, and specifically to elaborate on what is meant by the context of appreciation.

What the contexts of justification and systematisation share is that they describe the experience (of seeing something as so-and-so) in terms of seeing those things as having certain properties that function as grounds for a claim or assertion that that thing is so-and-so. In a context of appreciation, by contrast, we can have the experience of things being (perhaps) so-and-so (the 'perhaps' is important, as we shall see later), but without treating those things as being so-and-so in virtue of having such-and-such properties, and thus without relating with that thing as if it had such properties (that we could use as grounds for a claim or assertion). Put another way, the context of appreciation is different from the contexts of justification and systematisation in that appreciation is less assertive and less committal than justification and systematisation. In being less assertive and less committal, it is arguably much more relevant a cognitive ability for persons in social life understood from the second-person perspective.

To make the context of appreciation plausible we need to resist: 1) the idea that there is an underlying order of how things are, such that all we can learn are general rules that further constitute (or can constitute) a set or system, these being the rules that we then use to recognise how things are; 2) the idea that we necessarily, whether explicitly or implicitly, make claims or assertions about how things are that could be justified or systematised, i.e. that whenever we relate with the environment or others we are concerned about the potential status of our experiences (as if our experience was somehow inherently responsible, or necessarily committed to being responsible – instead of contingently disciplining itself to be responsible when needed); and 3) make plausible, on the basis of examples, the experience of seeing things as (perhaps) being so-and-so in a way that does not treat them as having such-and-such properties.

To help us to elaborate on appreciation, we shall enlist the help of Frank Sibley's work on aesthetics. We shall look – though again very selectively and narrowly – at aspects of two papers, 'Aesthetic and Non-Aesthetic' (1965) and 'Aesthetic Concepts' (1959), beginning with the former first. Having done so – and thereby hopefully made appreciation generally more plausible – we will then go on to consider, towards the end of this Chapter, how this general picture might apply to normative appreciation.

A. *Appreciating without Properties*

In 'Aesthetic and Non-Aesthetic' Sibley fastens onto what he calls 'aesthetic judgement'. It is important to see that this has a particular meaning for Sibley: for instance, it does not refer to 'verdicts', which are judgements that some 'things are aesthetically good or bad, excellent or mediocre, superior to others or inferior, and so on' (Sibley 1965, 136). Further, and importantly, it does not refer to judging that something 'is, say, large, circular, green, slow or monosyllabic', i.e. to put it generally, to judgements about 'the shape, colour, sound, wording, subject matter, or composition of things' (which Sibley characterises as non-aesthetic: 1965, 135). Rather, aesthetic judgement refers to the ability to 'see, notice, or otherwise perceive, for instance, that something is graceful, dainty, or garish, or that a work of art is balanced, moving, or powerful' (Sibley 1965, 135).

What is crucial here is how Sibley enables (indeed argues for) a focus on aesthetic judgement in his sense. He does this in a very instructive way. First, he says that the focus has to be on the ability 'to see the grace or unity of a work, hear the plaintiveness or frenzy in the music, notice the gaudiness of a colour scheme, feel the power of a novel, its mood, or its uncertainty of tone' (Sibley 1965, 137). The focus, then, must be on the seeing or feeling of a quality. Second, and in order to support this, he argues that it would not be sufficient to account for aesthetic judgement to focus on persons' ability to recognise that such-and-such a set of non-aesthetic properties (such-and-such a grouping of shape, colour, etc) would (potentially) justify a claim that something could be (perhaps should be, according to the person making the claim) experienced as being, for instance, gaudy or graceful. It is precisely Sibley's focus on seeing or feeling qualities such as gaudiness or gracefulness that is helpful for articulating the context of appreciation. It is helpful because it shows us that it would be implausible to say that every time we exercise aesthetic judgement it

would be necessary and sufficient for us to be guided by some general rules (constituting, say, our stock of general, unified knowledge of aesthetic beauty) that identify certain properties upon which certain qualities rely.[233]

We mentioned above that our focus on Sibley is selective, and here is the first reason why. Part of Sibley's discussion in 'Aesthetic and Non-Aesthetic' is on a critic's judgement that something is, say, gaudy or graceful. The problem, for our discussion, of focusing on the critic's judgement is that we would then tend to think precisely of the way in which the critic's judgement could be said to be 'responsible and trustworthy' (Sibley 1965, 145). Requiring judgement to be responsible and trustworthy is in itself not a problem – the problem lies in how one would read the requirement of responsibility and trustworthiness. If one reads it so as not to allow the person making an aesthetic appreciation to take a risk, i.e. to appreciate something as gaudy or delicate in a way that this person knows (for argument's sake) others are not likely to accept as gaudy or delicate, then this would be reading responsibility and trustworthiness in too demanding a way – a way that would have the effect of squeezing out the possibility of appreciation (the point being that appreciation is, always, somewhat risky: its correctness is not guaranteed, but subject to negotiation, potentially never being resolved one way or the other). As an argument can be had over the kind of pressure that a critic is under, especially when the critic is envisaged as having a public role that requires making (justifiable, systematisable) claims or assertions, what we propose is putting the image of the critic aside.

We need, then, to focus on everyday seeings or feelings of qualities that may, furthermore, be oriented towards not some objects that we might already readily classify as 'objects for aesthetic judgements' (such as paintings and the like), but everyday things, including 'scenery and sunsets, animals, faces, and people' (Sibley 1965, 136). What we need to take from Sibley, then, is the possibility of a focus on the everyday exercise of abilities to appreciate qualities that does not operate under the pressure (expected or demanded) of having to see or feel something as if it had properties that could be treated as the bases for a justifiable claim or assertion. What we need to focus on, then, is the ability to see or feel something could perhaps be graceful or gaudy, but without us treating (and being under implicit or explicit

---

[233] Sibley does not make the split we are making between contexts of justification, systematisation and appreciation. No claim is here made about the concept 'aesthetic judgement'; instead, we are using an analogy with Sibley's treatment of aesthetic judgement to make a case for the distinctiveness of appreciation.

pressure to treat) that thing as having certain properties that could justify us seeing or feeling it could be so.

Again: there is no doubt that we can see or feel something in the course of, and by way of, trying to provide some bases for a claim or assertion (communicated, or even only potentially communicable, to someone) as to its being classifiable as such-and-such (e.g. graceful or gaudy) on the basis of such-and-such properties, but there is no need – we do not have to be so anxious as to think there must be a need – for some kind of implicit self-discipline to only find those things graceful and gaudy that others might recognise as so on the basis of some rules (for when inference that things are so is warranted) that we all allegedly share.

There are some statements in 'Aesthetic and Non-Aesthetic' that may help us in this respect. Thus, for example, Sibley asserts that 'there are no sets of non-aesthetic features that are logically sufficient for it to have a certain aesthetic quality', i.e. 'There is no one set, no group of sets of logically sufficient conditions, no "defeasible" sets, and no wholly non-aesthetic descriptions which logically entail a certain aesthetic character or in virtue of which to deny such a character would be a linguistic error' (Sibley 1965, 152). Aesthetic qualities, adds Sibley, 'cannot be defined in a certain way solely by non-aesthetic terms' (Sibley 1965, 152). And again later: 'sometimes despite its fast tempo a piece of music has a sad quality, and sometimes despite a certain angularity in a dancer's movements she is nevertheless graceful' (Sibley 1965, 154). This last sentence is potentially misleading, suggesting defeasible application of general criteria (i.e. criteria that always has the status of being general and applicable, unless given reason otherwise). This is misleading, for though it may apply to the context of justification – in that context, it might make sense to speak of a critic having certain tools in a toolbox, i.e. certain markers he tends to give in support of certain aesthetic judgements – it need not be determinative of appreciation. The point can be put in a positive way: I do not need to, in order to appreciate some piece of music as sad, think to myself that 'it is sad *despite* being fast' – I can simply appreciate it as sad. In appreciating it as sad I do not, as it were, *attribute* a quality to it – my appreciating is not a claim or assertion that it is so – and I do not treat it as having a certain quality because (or despite) it has such-and-such properties (for in appreciating it I am not treating it as something that provides

grounds for my experience of it); I am simply appreciating it as something that could (perhaps) be sad.[234]

It may help here to see that on one level of generality someone (a critic, say) could say that some non-aesthetic property (slow tempo) is correlated with sadness, and we could think this is more or less justifiable (one can imagine a community in which only slow tempo music is played at events that tend to be experienced as sad, e.g. funerals). But notice, even within this imaginary community (but all the more outside it), that there are many other possible levels: e.g. slow tempo being correlated with, as Sibley puts it, 'solemnity, majesty, pensiveness, and serenity' (Sibley 1965, 156). Now the point here is not that there is an exhaustive catalogue of levels, and that we must, given an appreciation about it being slow, claim or assert that this piece of music is either sad or solemn, majestic, pensive or serene. Instead, the point is to see the limitations of the correlations (that they are but devices that may be used in contexts of justification and systematisation, but that they are not – because they can be useful in those contexts – determinative of the operation of appreciation).

In a way, what we might say is this: when we prioritise, or situate ourselves only within, the contexts of justification and systematisation (though especially the former) we may speak of sad-making properties (identifying slowness as such a property), but here, in the context of appreciation, it is better to say that slowness may accompany our appreciation of something as sad, but that it also may not (and when it does not, it makes no sense – is only misleading, and not helpful – to say that we appreciate it as sad *despite* it not being slow). The point here, once again, is to avoid postulating some general, unified or unifiable, underlying order that determines, by default, what we appreciate, unless we are given reason (to stop and reflect) and appreciate otherwise. There is no need to stop and reflect, for there is no underlying order that determines how we appreciate, or indeed any underlying order of properties constituting certain qualities – i.e. there is no determinate way in which we are such that we are enabled to recognise something as being in some specific way, and no

---

[234] I can work my back to characterise some properties if need be – but I may be constrained in my articulation of these in all kinds of ways (e.g. via the availability of terms, my language skills, the language skills of my audience if there is one, and so on). My *post facto* process ought not, then, to be understood as 'revealing' the bases upon which I make my appreciation: appreciation outruns, so to speak, articulation (though articulation can, in its own way, help appreciation – sometimes, I can learn to appreciate better by disciplining myself with subtle articulated distinctions, but no such articulation fully controls my exercise of appreciation).

determinate way in which the environment is (or someone in that environment is) such that it (or someone) must be recognised in some specific way.[235]

B. Taste, or the Ability to Appreciate

Such is as much as we can glean, for present purposes, from Sibley's 'Aesthetic and Non-Aesthetic'. Let us now turn, briefly, to the earlier 'Aesthetic Concepts' (1959). This paper helps us because it focuses on taste, and in that way assists us in seeing how we can speak of having abilities to appreciate (and not just, then, appreciating on isolated occasions). This is important because otherwise the advocate of the contexts of justification or systematisation could say that, certainly, there may be isolated instances of appreciating things in some way not previously appreciated, but this is not very significant: what is significant is that persons develop abilities to claim or assert that things are so-and-so, and the only way to conceive of these abilities is by prioritising the contexts of justification or systematisation (for what, if not that, could mean that someone has an ability?).

This issue returns us to something we have faced before in this thesis: namely, how to understand expertise. Focusing on taste is helpful, for it can assist in making room for an intuition we might otherwise have trouble making if we focused on expertise. Here, briefly, is why: we often imagine, when we speak of expertise, of a domain (the expert domain) being constituted by certain rules that determine what counts as what, such that an expert is then pictured as someone who has learnt and internalised ('mastered', as is often said) those rules to such an extent that they can now discern those properties that the rules (in their antecedent) require as the grounds for making certain claims or assertions (about things being so). It is this view of expertise that we are unravelling here. It helps to think of taste because it is intuitively less plausible to think that someone who has taste (and not necessarily a critic, but anyone) has learnt rules that, for instance, point to those properties that allegedly make an object graceful or gaudy, such that all that is required for someone to develop taste is to learn and internalise those rules, and then spot (without any reflective deliberation, as if they had a sensitivity to) those properties and on that basis make the claim or assertion that the object is graceful or gaudy. This is a less

---

[235] This does not mean that the other might not feel in some specific way and want us to notice that. The point is to avoid a description of the example that would force us to describe the relation of the person with another person as being responsive to the particular way in which we, as a theorist, describe that other person to be.

plausible picture partly because it seems implausible to think that what is graceful or gaudy is so rigidly rule-governed, i.e. that it is only on the basis of having such-and-such properties that something could be appreciated as being graceful or gaudy. Indeed, what we want to say here is the opposite: that someone who has taste is not someone who follows rules that tells them when something is gaudy or graceful (because they discern the allegedly gaudy-making or graceful-making properties), but precisely someone who, though she might have the ability to justify things (in a way she thinks likely to be accepted by other persons in some community) as gaudy or graceful because of such-and-such a property attributable to it, can appreciate things as being graceful or gaudy in novel ways that in fact create properties via which something can, in the future, be justifiably or in a systematic way, claimed or asserted to be graceful or gaudy. Those with taste are not those who have become supremely passive vis-à-vis alleged quality-making properties, but precisely those who have become active with respect to them (recognise, as it were, their ambivalence and openness).

Here we can pause to say something important about the context of systematisation. This context depends on taking a certain range of objects and characterising them (comparing and contrasting them) so as to articulate their differences, while perhaps also connecting them under a common category. The interesting thing to see about this is that if we treat that pool as stable – i.e. if we introduce no other objects, but just see the general properties used to make that pool a particular mixture of differences and similarities – then we also make it difficult to see how the objects could be anything but constituted by those properties. But what can happen – either by accident or by my introducing a new object – is that the pool can be enlarged, at which point how things relate to each other (how they are distinguished from each other, including how they may all be the same in some way) can be changed (of course, it need not be: a person can simply dismiss the object as not properly a member of the pool, but this is not determined in advance).

When we prioritise the context of systematisation, then, we tend to treat things to be in a certain well-established pool, where there are well-established ways in which things differ from each other while being the same in a certain way, and then we tend to classify whatever we are encountered with as either falling within this pool (and thus classifiable as both the same in some way, but also as an instance of a certain difference in the pool) or falling outside the pool. What the context of

appreciation reminds us is that this is but one context in which we can relate to the environment, i.e. we do not have to treat (just because it would make sense from the perspective of the context of systematisation to treat) that pool (that system of differences and similarities) as unchangeable. The point here is not that it is a default that we can change in some way; instead, the very fact that we can change it (by accident or by introducing on purpose an object that puts pressure on that way of systematising) should show us that it is but a context, and not something then that necessarily (by default) guides appreciation. Appreciation is its own context, and not one driven by – unless given reason to do otherwise – the context of systematisation.

Let us return to taste. How does Sibley talk of taste? Part of Sibley's discussion is not helpful to us because he describes the process of making an aesthetic judgement under the pressure of having to answer the question 'why?'. At issue (as with the question of responsibility and trustworthiness raised above) is how to understand what the 'why' question asks for, and how it can be answered. For example, it would be problematic from our perspective to say that whenever someone, whether a critic or not, exercises aesthetic judgement, they are under pressure to make all of their pronouncements justifiable, and also coherent (systematisable) with each other. On the view developed here, we would prefer to understand that pressure as a contingent one, and not present by necessity. If it were present by necessity, such an exercise of taste would be implicitly or explicitly a matter of answering why questions by identifying properties that allow for inferences to something being, e.g. delicate or gaudy. On our account, exercising taste involves being able to appreciate in ways that switches the burden, i.e. it puts pressure on the ways in which things may, at any one time, be said to be justified and coherent. However, as arguments can be had here concerning the role of a critic (for it may be claimed that it is the critic's role to be able to answer 'why' questions precisely with justifiable, systematisable, claims or assertions) we shall leave the matter aside. In other words, the argument is that we need not be critics (on this understanding of what being a critic means) to have and exercise taste.[236]

---

[236] In fact, this too is more complex that this paragraph suggests, but Sibley does not elaborate enough for us to analyse it properly. He seems to suggest that there is a way of answering why questions in aesthetics that is not the same as answering why questions in other domains: that though our answers may refer to reasons for which something is, say, delicate, they do so in a different way. Sibley asserts, for instance, that saying 'it is nobly austere because of the lack of detail and the restricted palette' is different to saying 'he is lazy because of the way he regularly leaves his work unfinished' (Sibley 1959, 435-6), but he does not articulate why he thinks the use is different. He does say that when we

Some things that Sibley says, however, are useful for articulating the ability to appreciate. For instance, Sibley says that 'Things may be described to us in non-aesthetic terms as fully as we please but we are not thereby put in the position of having to admit (or being unable to deny) that they are [for example] delicate or graceful or garish or exquisitely balanced' (Sibley 1959, 426). This is not the same as saying, as Sibley helpfully argues, that there are defeasible delicate-making or graceful-making properties that apply (and guide the person with taste) unless given reason to think otherwise: 'We could not conclude, even if we were told of the absence of all "voiding" or uncharacteristic features (no angularities, and the like), that an object must certainly be graceful, however fully it was described to us as possessing features characteristic of gracefulness' (Sibley 1959, 430).

This reminds us, as Sibley's 'Aesthetic and Non-Aesthetic' paper also reminded us, of the importance of seeing and / or feeling something as, e.g. delicate or graceful (it is not enough to simply hypothesise that something is likely to be found delicate or graceful if it has such-and-such properties: one has to experience it). The point here is that seeing and feeling cannot be described as if they were only ever going to confirm what we already use to justify or systematise claims or assertions about delicacy or gracefulness: to do that, as the defeasible-default structure does, would be precisely to impose an underlying order in ways that force us to describe ourselves as much more rigid and passive than we can be.

Sibley goes on to elaborate on what he means by a fascinating discussion of how to think about examples of aesthetic qualities. We need to quote him at length here, for it is a most instructive analysis. In the passage below, Sibley begins with concepts that are not aesthetic, such as 'intelligence' or 'laziness', and he says of them that they are more open and complex than his contrast with aesthetic qualities (in preceding paragraphs, not included here) suggests. But it is what he does next that is so interesting, for he then says that aesthetic concepts are even more open and complex than that, i.e. that we need to avoid thinking of them in the same way as we would of the other concepts (those that we can think of as having, as he puts it 'samples or precedents' that 'necessarily embody…the complex web of governing and relevant conditions'):

---

are offering aesthetic reasons, we are not offering 'explanatory or justifying conditions' (Sibley 1959, 435) but he does not say what we are offering instead. The difficulty here may be precisely that Sibley thinks we may need to offer something, our point being – by referring to appreciation – that we do not.

Many concepts [Sibley is referring here to aesthetic concepts], including most of the examples I have used (intelligent, and so on…), are much more thoroughly open and complex than my illustrations suggest. Not only may there be an open list of relevant conditions; it may be impossible to give rules telling how many features form the list are needed for a sufficient set or in which combinations; impossible similarly to give rules covering the extent or degree to which such features need to be present in those combinations. Indeed, we may have to abandon as futile any attempt to describe conditions or formulate rules, and content ourselves with giving only some very general account of the concept, making reference to samples or cases or precedents. We cannot master or employ these concepts therefore simply by being equipped with lists of conditions, readily applicable procedures or sets of rules, however complex. For to exhibit a mastery of one of these concepts we must be able to go ahead and apply the word correctly to new individual cases, at least to central ones; and each new case may be a uniquely different object, just as each intelligent child or student may differ from others in relevant features and exhibit a unique combination of kinds and degrees of achievement and ability. In dealing with these new cases mechanical rules and procedures would be useless; we have to exercise our judgment, guided by a complex set of examples and precedents. Here then there is a marked superficial similarity to aesthetic concepts. For in using aesthetic terms we learn from samples and examples, not rules, and we have to apply them, likewise, without guidance by rules or readily applicable procedures, to new and unique instances. Neither kind of concept admits of a 'mechanical' employment.

Nevertheless it is at least noteworthy that in applying words like 'lazy' or 'intelligent' to new and unique instances we say that we are required to exercise judgment; it would be indeed odd to say that we are exercising taste. In exercising judgment we are called upon to weigh the pros and cons against each other, and perhaps sometimes to decide whether a quite new feature is to be counted as weighing on one side or on the other. But this goes to show that, though we may learn from and rely upon samples and precedents rather than a set of stated conditions, we are not out of the realm of general conditions and guiding principles. Samples and precedents necessarily embody, and are used by us to illustrate, the complex web of governing and relevant conditions. To profit by precedents we have to understand them; and we must argue consistently from case to case. This is the very function of precedents. Thus it is possible, even with these very loosely condition-governed concepts, to take clear or paradigm cases, to say 'this is X because…' and follow it up with an account of features which clinch the matter.

Nothing like this is possible with aesthetic terms. Examples undoubtedly play a crucial role in giving us a grasp of these concepts; but we do not and cannot derive from these examples conditions and principles, however complex, which will guide us consistently and intelligibly in applying the terms to new cases. When, with a clear case of something which is in fact graceful or balanced or tightly-knit but which I have not seen, someone tells me why it is, what features make it so, it is always possible for me to wonder whether, in spite of these features, it really is graceful, balanced, and so on.

(Sibley 1959, 430-432)

What is fascinating in this passage is precisely Sibley's contrast between the exercise of judgement and the exercise of taste. When we think of taste, we cannot think of it as guided, even by defeasible default, by certain rules.[237] Indeed, in this respect, Sibley is very explicit: 'if someone did merely follow a rule we should not say he was exercising taste, and we should hesitate to admit that he had any real notion of delicacy until he satisfied us that he could discern it in other instances where no rule was available' (Sibley 1959, 433). What Sibley suggests, then, is that only someone 'lacking in [aesthetic] sensitivity' would need to learn to follow rules (see Sibley 1959, 433, fn. 6). The issue, Sibley further adds, is not one of specification: it is not that, say, there really is an underlying order of delicate-making properties, but we simply have not matured enough – because, say, our instruments are not precise enough – to recognise (and articulate) what they are: 'it is obvious that even with the help of precise names, or even samples and illustrations, of particular shades of colour, contours and lines, any attempt to state conditions would be futile…[because] the very same feature, say a colour of shape or line of a particular sort, which helps make one work may quite spoil another' (Sibley 1959, 434). Rather, 'it is a characteristic and essential feature of judgements which employ an aesthetic term that they cannot be made by appealing…to conditions…. This is what taste means' (Sibley 1959, 436-7).

There is a lot to take from this in clarifying appreciation, but some also to be resisted. For instance, we should resist the idea of there being different kinds of concepts; some that are amenable to judgement and some to taste.[238] Instead, we should speak – as we have been here – in terms of contexts in which we might treat concepts (*all* concepts) as being ones either governed by rules (this we do in contexts of justification or systematisation – where nevertheless, as Sibley points out, the concepts still remain open and complex to some extent, meaning that we are still active and creative in certain ways in those contexts) or where we are not under

---

[237] Sibley characterises judgement by precedent (including in this the exercise of legal judgement) in a certain way to bring out certain features of what he wants to present as taste. What he says about taste is very instructive, but this does not mean we need to endorse what he says about (legal) judgement by precedent. For a reading of the practice of (common law) precedent different to Sibley's, see Del Mar 2011c.

[238] We are, of course, referring here to Sibley's terms; in our terms it would be 'contexts of justification and systematisation' (instead of Sibley's 'judgement') and 'context of appreciation' (instead of Sibley's 'taste').

guidance or under pressure to be guided by any rules (precisely the context of appreciation). The fact that we are not so guided, however, in the context of appreciation does not mean that we are not having the experience of something as (perhaps) so-and-so, e.g. perhaps delicate, perhaps graceful; indeed, the point is that we are *both* not being guided *and* that we are having such an experience.

What Sibley's discussion of taste adds is that we can see this experience of things as (perhaps) being so-and-so (in the context of appreciation) as an ability, as something we develop and can become good at. We can develop abilities to appreciate (that are not restricted to certain kinds of concepts), and these are precisely abilities that depend on there not being, and not being guided by, some general, unified or unifiable structure, determining what counts as what (identifying properties that make objects have certain qualities).

Of course, this does not mean that before we can appreciate anything (e.g. delicacy) we must have an ability to do so – to say that would be, in effect, to offer another version of a deterministic explanation. We can appreciate something as delicate for the first time. One might ask here: but how do we know when we appreciate it that we are appreciating delicacy, and not something else? The answer is that we do not: we could try and explain the experience to someone, and they might tell us that this resembles 'delicacy', but we might also invent a new mode of appreciation and give it a new name – of course, we may also call it 'delicacy' to appease our friend, but nevertheless go on appreciating it in our singular way.[239]

But, as Sibley's discussion of taste hopefully helps us see, we can also develop an ability to appreciate delicacy: we can become good at appreciating delicacy. This does not mean that we can make claims or assertions that we can justify or systematise about delicacy as a quality – the point here is not to say that by becoming good at appreciating delicacy we can lay claim to being a critic (on that way of understanding what a critic does); instead, the point is that justifying or systematising claims or assertions about delicacy is not necessary for, nor in some sense prior to or

---

[239] We do not wish to challenge the idea that part of the normative dimension of social life may involve deference to 'ways of speaking', or ways of articulating how things may be justified and systematised (in certain domains, or certain communities). Further, we acknowledge that there may be contexts where certain resources for justification and systematisation may be required to be used (e.g. judges in common law courts are required to articulate their judgement by reference to relevant past decisions of higher courts – but see Del Mar 2011c for a reading of this exercise of legal judgement). But acknowledging all of this does not threaten the basic argument: that the underlying ability is the ability to appreciate, and that this ability is always somewhat independent of any normative resources that may be utilised in the contexts of justification and systematisation.

determinative of, appreciating delicacy for the first time or becoming good at appreciating delicacy.

C. The Tentativeness of Appreciation

Throughout this Chapter we have been signalling a kind of tentativeness and hesitancy at the core of appreciation by saying that when we appreciate we appreciate things as '(perhaps) being so'; that we appreciate that something 'could be so'. This is important because it reveals an ongoing openness that is characteristic of appreciation. When I appreciate, precisely because I do not need to claim or assert it is so, I leave open the possibility that, e.g. some piece of music might be sad, but that it might not, and in 'might not', we mean both that it might be better appreciated as happy (and thus in a sense opposite to sad) or that it might be better appreciated as sad in some more particular way, e.g. sad in an ironic sense, or sad in a nostalgic sense (or sad in a sense we need a new name for).

Appreciating is not a stop-start business in which we can make one claim after another (even if the next claim defeats the previous one); it is inherently more tentative and hesitant – it does not commit us to one way of appreciating something. We continue appreciating as we continue to relate with whatever it is that we are relating with. Another way to put this is that whatever I am relating with, e.g. a piece of music, invites me to, or offers me opportunities to, appreciate it in ways that are not closed off – that remain open as I go on appreciating it.

This sense of hesitation or tentativeness (of appreciation as an ongoing mode of exploration of how things might be) is important, for without it we might think that appreciation is solipsistic: that I am somehow locked into appreciating in the way I do and cannot, for instance, be corrected (that I would not be open to being corrected but rather that I would stamp my foot and say: but that is just how it appears to me). When I appreciate, I am inherently open to correction – that is precisely what is meant by my being hesitant and tentative; I do not need to commit myself to one of appreciating it in order to be open to correction in the course of relating with it. When we say 'correcting' here, we do not mean 'correcting' in accordance with some scheme, but just that, as I go on relating, I am open to being given or indeed to creating a reason to appreciate it differently, as when I appreciate some music as sad but then go on, while listening to it, to thinking it is sad in a more particular way

(perhaps in a way I have not appreciated things as being sad before – which might, in fact, turn out to be more satisfactorily appreciated as not sad at all).

This is connected to why we have to see that abilities to appreciate are not describable as requiring or facilitating immediacy or directness, or a sense of phenomenological ease and fluency. To describe the exercise of abilities to appreciate in that way would be misleading, for it would retain the commitment to an underlying order and make us think of our abilities as the result of internalised rules that constitute that order. To avoid these consequences of the image of immediacy, we need to see that it is more accurate to describe these abilities as being exercised in a tentative way, as when I appreciate that something 'could be' or is 'perhaps' graceful, delicate or gaudy. This also means that the process of appreciation can be difficult, not easy, and a mix of confidence and doubt, rather than some kind of assertive fluency.

When I have an ability to appreciate, I am more, and not less, active, creative, open and flexible, and I am more, and not less, willing to explore in a tentative way, experimenting with ways of appreciating things. It is only when we think there is an underlying order that we require the person to have a sensitivity that enables her to immediately recognise things as being so-and-so (in accordance with the rules of the underlying order). For appreciation, there is no scheme in the background in accordance with which things must be appreciated as being so-and-so: indeed, here we see how appreciation itself is only possible if one allows for the environment, and also us, to be at bottom indeterminate (the point being there are no fixed so-and-so's waiting to be recognised as being so-and-so).

We have spoken a good deal here about appreciation in the context of person-to-environment relations (on the back of the analogy with aesthetic judgement), and the connection with interaction (from the second-person perspective – as elaborated upon in Part V) may not be clear. When we interact, you may exercise your ability to appreciate in various ways. For instance, you may appreciate me as (perhaps) being in such-and-such a way (e.g. appreciating my movements as elegant). Or, you in a more actively interactive sense, you may appreciate me as offering (perhaps) such-and-such an invitation or opportunity to relate, e.g. you may appreciate me as offering my hand for a greeting gesture (imagine that you have never before shaken hands when greeting: appreciation is what allows you to be creative in what you appreciate as a greeting), or you may appreciate moving my hand as part of making a compliment or

issuing a threat. The point of appreciation is that you can appreciate my movements as, e.g. a compliment – calling perhaps for a thank you, or for a compliment in return, or something else you have never before done in answer to a compliment – without the need to apply some general schema, and upon that basis convince yourself I am really complimenting you; instead, your ability to appreciate – in this case, appreciate in a more interactive way – enables you to see my hand as calling in an unfamiliar way for a response that may be unfamiliar to you.

Appreciation, then, may be understood in more or less actively interactive ways. However it is understood, what is vital to see is that it offers a mode of relating cognitively with the other that is more tentative and hesitant than the making of a (justifiable, systematisable) claim or assertion. It is, because of that, also more risky – it does not involve, like the making of a justifiable or systematisable claim involves, appealing or being concerned about the status of one's experience of the other. It does not attempt to be as responsible an experience as the experience one has when making a justifiable or systematisable claim. And yet, it is a form of experience that is richly engaged with the other – it is not solipsistic and it is inherently open to revision. It is that which allows us to be surprised, and respond, to what others do and say in ways that are new to us, and do not depend on the prior acquisition and / or internalisation of rules. Because of all this, appreciation is precisely the cognitive ability that persons need to participate in social life from the second-person perspective.

D. Normative Appreciation

Like appreciation in general, normative appreciation is also a tentative, hesitant and inherently risky and revisable experience of another (and possibly the environment) as perhaps being so-and-so. For instance, just as I can tentatively experience an artwork as perhaps elegant or a piece of music as perhaps sad, so I can tentatively experience another as having done something perhaps inappropriate, with perhaps such-and-such a response called for. Or, I can tentatively experience another as perhaps issuing an invitation or opportunity for me to respond in such-and-such a way, where so responding can be experienced as perhaps the appropriate thing to do in the course of interacting with that other.

Already, however, in the above paragraph, we may begin to see some of the differences between appreciation in general and normative appreciation. The difference lies in the connection with action. Whereas general appreciation is more

amenable to being undertaken without going on to do anything specific, normative appreciation is more likely to involve the emergence of a will to do something, as a consequence of making that appreciation. In this respect, normative appreciation shares something similar with the kind of justifiable or systematisable claim we can make on the basis of some already known normative resource. In other words, when I make a claim that someone has done something inappropriate, e.g. I evaluate someone as having breached such-and-such a norm (e.g. speaking loudly on a mobile phone in a bus), then this is likely to involve some reaction – for instance, asking or at least looking for a justification for that person's behaviour (e.g. one looks for a reason for the level of volume that person is speaking – and sees that the person has a hearing aid). This is unlike a general appreciation, as when I appreciate some artwork as elegant, no reaction need follow, unless perhaps I am disposed to purchasing artworks that I appreciate as elegant. Normative appreciation, then, is more closely connected with action than appreciation in general.

Nevertheless, it is important to see that normative appreciation need not be tied to any particular kind of action. In other words, just in the same way in which I tentatively appreciate something as perhaps appropriate or inappropriate, so I can tentatively appreciate doing something in response as perhaps appropriate or inappropriate. I have, there, two kinds of normative experiences – my appreciation of what someone does, and my appreciation of what I do in reply – both of which are tentative and hesitant, and indeed very highly revisable. Indeed, the advantage of normative appreciation – as opposed to normative justification or systematisation – is that my appreciation both of what someone does and of what I ought to do in reply are highly sensitive to change, or differently put, to seeking cues for either positive or negative reactions to how I have appreciated the appropriateness of what someone did and what I have done (or am doing) in reply, and revising what I do, and what I take myself to be doing, in light of those cues. All of these appreciations, it must again be emphasised, are tentative and hesitant – they are forms of involved, engaged observance that are minimally assertive and minimally committal, and yet still outward-looking (not solipsistic) and connected to action. They are, furthermore, not dependent on any previously acquired knowledge about how to appreciate someone as appropriate in this context, or how best to reply to someone when they are so appreciated. They are, in that sense, risky: they may be experienced – appreciated – by the other person in the interaction as inappropriate. But their being risky is what

makes them useful; indeed, their being risky is what allows persons to learn and adapt and interact with persons who are wholly different to them.

Part of the point in making normative appreciation so tentative and hesitant is that it does not fall into the trap of holding in place – pre-determining – what is normatively relevant in a particular situation. In other words, when we describe how persons appreciate normatively, we avoid setting up the case in such a way where there is but one way of appreciating that something as appropriate or inappropriate. If we did this, we would be positioning persons very passively, and we would be missing that aspect of persons' cognitive abilities that are more active (paradoxically, perhaps, because they are less assertive and less committal). Arguably, this also helps us to be more accurate in our descriptions of how persons interact: after all, when we interact with another person we are not sure (and indeed, if we are sure we are likely to be very unsuccessful at participating in social life) what that other person might find appropriate or inappropriate. Certainly, prior interactions with another person might make us more confident, and we might come to rely on certain ways in which we have interacted – including shared judgements about this or that – with that person. But much of social life – perhaps especially in modern conditions of sprawling urban cities – is interaction with strangers, and sometimes in contexts that are unfamiliar to us. More generally, though, even in interactions with persons we know well, and even in familiar contexts, our interactions with others – as understood from the second-person perspective – involve a mix or balance of the familiar and unfamiliar, the predictable and unpredictable, the known and the unknown. In other words, even when we are relying on normative resources, and thus exhibiting the attitudes we discussed in Part I (as with conventions and norms) or assuming certain rules as relevant, we are still also, even to a small degree, reliant on our abilities to appreciate. We must remain, then, always somewhat tentative and hesitant in our relations with others, at least when we understand those relations from the second-person perspective.

Although we cannot develop it any detail here, it might be helpful to mention that the further development of this notion of normative appreciation might be aided by looking at certain strands in the moral philosophical and moral psychological literature. For instance, the call, by Jonathon Dancy, for a 'bottom-up moral

epistemology' (see Dancy 2010, in response to Audi 2010), might be drawn on.[240] Other possible resources might include the burgeoning literature on moral perception[241] and the so-called skill model of moral judgement.[242] All of these may serve as useful resources for further developing the notion of normative appreciation because they distance themselves from arguably over-intellectualised accounts that require conscious, deliberative, typically concept-infused reasoning on the basis of previously acquired rules – and, in that sense, may be more amenable to a view that describes normative experiences as tentative and hesitant (including seeing the value – perhaps even for moral life – of tentativeness and hesitancy). Notice, in that respect, that whether the account requires conscious deliberation or not is a red herring. The point is not to make normative appreciation easy, fluent and familiar; appreciation can be more or less deliberative, and is arguably always to some extent difficult. The point, instead, is that normative appreciation is less prepared – and thus less pre-determined – than a judgement that makes a justifiable or systematisable claim. It is, in that sense, less intellectual.

One final, and perhaps most promising, of resources could be that of the literature on emotions, especially where emotions are approached as appraisals. This approach is interesting, too, because of the link recognised between emotion and action. Thus, for appraisal theories, 'emotions consist of patterns of perception, or rather interpretation' (Ellsworth and Scherer 2003, 572). But perception (or interpretation) – or 'understanding a situation' (phrases all used by Ellsworth and Scherer) – is not enough: an organism must also be motivated to do something about it.[243] What is perhaps most important from our perspective is that emotions – so conceived in appraisal theories – are thought to be *flexible* ways of achieving a connection between perception and action: unlike mere 'mechanism[s] that trigger fixed actions patterns in response to appropriate stimuli', emotions 'respond to stimulus patterns with behaviour' while allowing for 'reconsideration' (Ellsworth and

---

[240] The extent to which it can be will depend on whether Dancy makes room for an experience of something as right or wrong (to use his language) that can create what might (or indeed might – hence the risk) be treated as a right-making or wrong-making feature.

[241] This literature may be helpful precisely because it attempts to move beyond the dichotomy of intelligent belief and dumb perception in the context of normative experiences. See, for instance: Cullison 2010; Chappell 2008; McGrath 2004; McBrayer 2010a and 2010b; Watkins and Jolley 2002; and Starkey 2006.

[242] See, for example: Goldie 2007 and Jacobsen 2005.

[243] For more on this link between emotional perception of values and action, see .g. Johnston 2001; de Sousa 1987 and 2002; Tappolet 2000; Wedgewood 2001; Mulligan 1998; Brady 2010; Deonna 2006; Döring 2007; D'Arms and Jacobson 2006; for a recent criticism of these views, see Whiting 2012.

Scherer 2003, 572). For example, 'fear creates a tendency to flee, but a person may quickly realise that the threat is directed at someone else (reinterpretation of the event) or that an aggressive stance will intimidate the attacked (reinterpretation of response alternatives)' (Ellsworth and Scherer 2003, 572). The key, then, as to whether appraisal theories of emotions will be useful for the further development of the notion of normative appreciation is whether their important linkage of perception (or interpretation / understanding) and action recognises a mix, and perhaps a degree, of doubt and confidence in that link.[244] If these theories can build in degrees of flexibility, and avoid rigid associations of one perception (or interpretation / understanding) with a specified action, thereby making room for tentativeness and hesitancy as well as the emergence of new associations between certain perceptions and certain actions, then they will not only be great resources for normative appreciation – they will also make a lasting contribution to our understanding of the abilities we need to participate in social life, including its normative dimension, as understood from the second-person perspective.[245]

---

[244] For instance, some of the relevant literature here concerns enactive approaches to emotions, as in the 'enactive appraisal' approach of Colombetti 2007. Colombetti places emphasis on the role of the body and understands that role via 'dynamical systems theory'. On this model, actions are conceptualised 'as emergent from the ongoing reciprocal interactions (or, better, "couplings") of embodied agents and their environment' (Colombetti 2007, 541). This reference to 'ongoing reciprocal interactions' is promising, but what is key is whether the approach is able to resist describing an agent's interactions with the environment as immediate. The advantage, then, of placing more emphasis on the body ought to be making room for more explorative kinds of relations, one's that do not require our relation with the environment to be immediately meaningful in some specific way.

[245] In this respect, Julien Deonna and Fabrice Teroni's recent work on the emotions might also be helpful, especially in their endorsement of an account of emotions as 'intimately connected with…felt action readiness' (Deonna and Teroni 2012, 79l; and see chapter 7 more generally). The world-directed, felt-bodily attitude of action readiness could be a way of further cashing out the notion of normative appreciation (thereby capturing its emotional and corporeal dimensions), but only as long as sufficient tentativeness and hesitancy is built into the concept of 'readiness'.

Chapter 25. Attentive Silence

The exercise of appreciation is of vital importance for the picture of cognition being developed here, i.e. precisely the picture of cognition that emerges from being situated in interaction from the second-person perspective. Of importance to this picture, aside from the exercise of appreciation (which was considered in the preceding Chapter), and in partnership with it, is what is called here 'attentive silence'.

The phrase 'the silence of the senses' is the title of a paper by Charles Travis (2004), and we will be looking at this paper in a moment. For now, let us motivate the idea with an example: imagine you are running your hand over an unfamiliar piece of material. You may be a tailor, but even if you are, you are not running your hand over it in the context of some tasks according to which you are obliged to give some report on the utility of the material for some dress; you are simply running your hand over the material. As you do so, it is possible that you appreciate it as being (perhaps) rough, or rough and smooth in parts, or rough-and-smooth in a way not yet previously appreciated by you; it is also possible that you run your hand over the material with a view to making a (justifiable or systematisable) claim or assertion about the material (about how it is); but although these are possibilities, they are not necessities: it is also possible that as you run your hand over the material, you are attending to it, relating with it, and not yet either appreciating it or making justifiable or systematisable assertions or claims about it. One could say that in this instance your ability to appreciate or make (justifiable or systematisable) claims or assertions is hovering above your hand, which is silently attending to the material – your hand is, as it were, inviting the material to affect it, to relate with it; it is not committing you to any way of feeling how it is or might be, but nevertheless, despite you not feeling how it is or might be, you are attending to it, relating with it.

Perhaps the best contrast with this notion of 'attentive silence' is daydreaming: I could be running my hand over the material and be recalling an episode in which I gave a bad lecture; then, my senses are silent in a way that is non-attentive – they are not actively allowing themselves to be affected (though affected they might still be, e.g. if there is a needle in the material that pricks my finger). Another, even more dramatic contrast, is with sleepwalking: I can be running my hand over the material while sleepwalking, but here it might be said that I am neither attentive to, nor even

silently non-attentive – it just does not seem as if (unless something really drastic were to happen that would jolt me out of my sleepwalking trance) the material could really affect me.

Attentive silence, then, is active, but non-assertive and non-committal – or, if we speak in degrees, it is less assertive and less committal than appreciation. It is not a matter of 'considering' the options of how something is or might be, but it is open or ready to appreciate or claim / assert. This does not mean, however, that it is attending only with a view to appreciating or claiming / asserting: although that is possible, it would be a very strong version of attentive silence – the basic case we want is one of attending silently in a way that may, but also may not, lead to appreciation or a claim / assertion.

A. *The Silence of the Senses*

We mentioned Travis's paper above. Travis's notion of 'the silence of the senses' takes its cue from John Austin's comments concerning the ways in which the senses could be 'dumb': 'In fact', Austin said in his *Sense and Sensibilia* (1962), 'our senses are dumb – though Descartes and others speak of "the testimony of the senses", our senses do not tell us anything, true or false' (Travis 2004, 64; see Austin 1962, 11). Travis understands this notion of the senses being 'dumb' in the following way: 'our senses merely bring our surroundings into view; afford us some sort of awareness of them. It is then for us to make of what is in our view what we can, or do' (Travis 2004, 64). It is important to see that Travis wants to avoid the view according to which the senses represent things as being so: the very point of positing the silence of the senses is that they can be seen as doing something, i.e. bringing things into view, but not in a way that requires taking it that X is so (this 'bringing things into view' is also what Travis calls 'unmediated awareness': Travis 2004, 65). But now the crucial move comes next, in terms of how Travis describes and characterises this 'bringing into view'. Consider the following passage:

> Another part of the point is that *perception*, as such, simply places our surroundings in view; affords us awareness of them. There is no commitment to their *being* one way or another. It confronts us with what is there, so that, by attending, noting, recognising, and otherwise exercising what capacities we have, we may, in some respect or other, make out what is there for what it is – or, again, fail to. It makes us aware, to some extent, of things (around us) being as they are. It is then up to us to make out, or try to, which particular

ways that is. Perception cannot present things as being other than they are. It cannot present some way things are not as what is so. That would not be mere confrontation. So it cannot represent anything as so. Representing, by nature, is liable to be of what is not so.

Such is a view, and, so far, only that. On it, in perception things re not presented, or represented, us as being thus and so. They are just presented to us, full stop. It is in making out, or trying to, what it is that we confront that we take things, rightly or wrongly, to be thus and so.
(Travis 2004, 65; original emphasis)

What we see here is that Travis nevertheless imagines that we are presented with some content: he does not call it 'representational' content, or the 'particular' taking of things to be so – but he nevertheless describes it as having an 'it'. This is partly because he imagines us, and describes us, as passively confronting that which is there, things 'being as they are'. In other words, in positioning us the way he does, faced with a world that is (imagined by Travis) in a certain way, Travis is forced to say that our senses, though they do not commit us to any 'particular' way of 'taking' how X is, do nevertheless present us with some X. Whatever he calls it (non-representational, non-particular-taking), Travis cannot get away from the picture in which 1) the world is in a certain way, such that 2) we (our senses) confront it in a certain way (the way it is), such that 3) we are then able to use our mental faculties to interpret some specific way in which the 'that' that we are confronted with (or presented with) is (or may be).

To see the trouble that this causes consider the following example:

Sid shows up for dinner drunk. That much we can all see. That he is drunk (at this hour) may mean that it will be a long and boring evening. Or, again, it may mean that, once again, he has lost his job. If it does mean that, and we are aware that it is does, then from what we see – that he is drunk – we may gain mediated awareness of those further facts. (Travis 2004, 66)

The senses here are said to present us with Sid being drunk – this is then 'the X' that forms the basis for interpretations or inferences. According to Travis, Sid being presented to us as drunk has 'factive meaning', which he argues is different to representation: it does not represent things as being so, but enables or facilitates interpretations or inferences (Travis 2004, 66). To say this, however, is but to extend the context of justification or systematisation: we picture a person as faced with some irreproachable nugget of fact, something pre-prepared for us (by our senses presenting us with what there is) and ripe for our (correct or incorrect) inferences as to what it

means (e.g. how it can be accurately explained, or accurately accounted for as having come to be the way it is). But now consider this: when we say 'Sid is drunk' we are already, surely, attributing a meaning to Sid's appearing to us in a certain way – either we are claiming or asserting Sid to be drunk, or we are appreciating Sid as (perhaps) drunk.[246] To see this, consider that whereas one of us may see Sid as drunk, another (perhaps she is a pharmacist) sees Sid as (possibly) under the influence of some powerful medication; yet another (perhaps he is someone who works with disabled persons) sees Sid as possibly having some kind of mental disability. Travis treats 'Sid being drunk' as something incontrovertible, and he does so for the purpose of comparing the correctness of different interpretations of why Sid is drunk: perhaps he was fired, or perhaps he happened upon 'a vodka tasting, which, atypically, he failed to resist' (Travis 2004, 67). The question Travis is interested in is what kinds of inferences may be warranted from X, i.e. the 'factive meaning' of Sid appearing (being presented to us by the senses) as drunk. But this is precisely to think that the senses are subordinate to (and dumb in comparison to) the interpretive (said to be complex or intelligent) tasks of the mind.

What we see here is that we make the split, and also give it status – i.e. make the split between the senses and the mind, and characterise the senses as dumb and the mind as intelligent[247] – when we fail to see alternatives, i.e. we describe the scene as one in which Sid really is drunk, such that our senses present us with this 'fact'. The reason why this is hard to resist is that when I am comparing what I take to be two alternative interpretations – reasons for why Sid is drunk – I am treating Sid being drunk as unchangeable, and so we conclude from this that it must be possible for the senses to somehow present us with the fact that Sid is drunk (for we think, how else could there to be two interpretations – they must be interpretations of some 'one' thing). In other words, we make an over-generalisation from a particular attitude we can take to something, i.e. an attitude of treating something as unchangeable, as when we use it as a standard for comparing things we take to be interpretations of it.

Now there is one popular reply to the idea that the senses are dumb (and thus a way to try to rebuff Travis's split between the senses and interpretations): it is to say

---

[246] The difference is that in the case of claiming or asserting Sid is drunk we would have in mind properties according to which we can justify that claim or assertion, whereas in appreciating Sid is drunk we can create what others may (but of course also may not) go on to treat as properties that may sometimes warrant an inference to someone being claimed or asserted to be drunk.

[247] We noted, above, that a similar kind of split sometimes characterises the opposition of belief and perception.

they are intelligent, but in a way that is necessarily immediate or direct, i.e. that the senses can be trained so as to immediately give us access to the fact that not only Sid is drunk, but also that Sid is drunk because he got fired (here one needs to see that on some versions seeing 'Sid is drunk' would be already something 'intelligent'). This is not a reply that we want to endorse. In other words, we want to resist this appeal to immediacy or directness; indeed, the possibility of attentive silence depends on there not being immediacy or directness (on Sid not appearing to as, necessarily and by default, in some meaningful way, e.g. as drunk, or as disabled, or as drunk-because-fired etc). To make the senses 'intelligent' in this 'immediate' way would be too quick, and would imprison us in the ways in which things appear to us (in this allegedly basic sense of 'factive meaning').

What we want to make room for, instead, is the way in which (recall the image of the hand running across the fabric) my eyes, ears and nose (and thus cross-modally) can be relating with how the environment (in this case Sid) appears to me, but without as yet (or even without necessarily desiring to, without the aim of) appreciating Sid as being (perhaps) drunk or judging (claiming or asserting) that Sid is drunk (e.g. doing so in a whisper to one's partner at the party). As a result of how Travis positions us – again, in an environment that is in a particular way (Sid is drunk), us seeing that (Sid is drunk) and then interpreting why (because he got fired, or some other possibility) – Travis misses out on both the phenomenon of attentive silence and appreciation, and imposes the structure of justification or systematisation, i.e. precisely seeing things, in this alleged basic way, as potential grounds on the basis of which one make more or less valid / correct inferences.[248]

Let us return to a phrase that Travis uses above, namely that the senses 'bring things into view'. This could be a helpful phrase, but it is important to avoid a certain way of understanding it. The way to avoid, which seems to be Travis's understanding, is one where one characterises the senses as presenting 'things' to us where 'things' are understood in a certain way (e.g. Sid appearing drunk). We want to avoid this because we do want to set up the case in such a way that there is something determinate in the environment for the agent to recognise; we want, instead, to set the case up in such a way that leaves open the way 'things' could appear to the agent. It is

---

[248] This critique can be further supported by the fact that Travis presents factive meaning as proof: 'That A factively means B, once recognised, makes A proof of B' (Travis 2004, 66). Our point here is that we are not always scanning the environment for proof.

crucial here, then, where one positions the agent. In order to make room for a second-person perspective, one must not position the agent in a scene that is (from the point of view of the theorist) imagined to be in a certain way, such that, then, the agent is also said – e.g. his or her senses are said – to be determined in a certain way, i.e. to see what there is allegedly to be seen. We see here, again, how the very possibility and plausibility of phenomena like attentive silence is dependent on certain methods of description, or more broadly, on certain ways of setting up cases / examples.

When, then, Travis says that our senses can 'bring things into view', or 'afford some awareness of them', this is helpful, but only as long as we do not imagine 'things' or 'them' to be in a certain way. When we say 'we' here we do not mean the agents in the scene, but 'we' the theorists describing it: the whole point is that we ourselves, as theorists, cannot impose a scheme of intelligibility on the acting agent, for as soon as we do so we are forced to characterise the cognition of that agent in overly rigid and passive ways (e.g. as being able to see, and only see, that Sid is drunk – the 'activity' of the agent then is limited to interpretations). There is a sense, then, in which we might say (following Travis) that attentive silence 'affords' the possibility of appreciating how things might be or claiming / asserting that they are so-and-so, but only if we preface this by saying that attentive silence is not necessarily followed by – that its point is not to prepare us for – acts of appreciation or claims / assertions. To say this we need to acknowledge that to attend silently is not to experience the environment as a 'fact' or a 'ground' for a warrantable claim or assertion. The absence of such a justificatory relation does not mean that we are not actively relating with the environment; on the contrary, we are being active but just in a non-committal and non-assertive way. Attentive silence, then, captures the kind of cognition we need when the agent experiences, tentatively and hesitantly, and yet actively and richly, the environment (or another person), as when, to return to our running example, as a tailor we run our hand over some fabric, relating with it by attending to it silently.

It was mentioned above that Travis not only misses attentive silence, but also appreciation, and one might ask why: does he not, after all, speak of us taking things to be so, and is not taking things to be so part of what appreciation is about? Yes, there is a sense in which appreciation is about appreciating things to be (perhaps) so, but it is important to see that this does equate to treating something as a ground (or proof) for an assertion or claim about how things are. To so equate it would be precisely to impose the contexts of justification or systematisation in a way that

squeezes out the possibility of appreciation. To appreciate is not to treat something as a ground (as something having certain properties that warrant a certain inference). The point is not to picture our cognition as necessarily and always under the pressure of scanning for proof or being passively confronted with proof. I can attend silently to Sid appearing to me, and Sid may appear to me: 1) without me going on to appreciate him as appearing to me as (perhaps) drunk, or disabled, or under the influence of medication (or anything else, including a new way in which I might appreciate him as perhaps being); and 2) without me going on to make (justifiable or systematisable) judgements about Sid (those two possibilities are just that: possibilities and not necessities). When I silently attend to the environment, I am actively, and in an embodied, sensory way, exploring, attending to, relating with the environment, allowing myself to possibly go on to appreciate or make a claim or assertion about it.

B. Seeking, Scrutinising and Seeing

We have mentioned several times already how language can deceive us here, for instance precisely in us having to say 'things' or 'something' being there – an X, or some scene with objects – as if it was there, in some way, ready to be interpreted (and ready, too, to serve as the standard for rival interpretations). One source that may help us with the difficulties in expression is Sibley's 'Seeking, Scrutinising and Seeing' (1955). The paper is useful because Sibley points out (by reference to a close reading of Ryle's *The Concept of Mind*, 1949), how often we use (and thus also how often we get misled by the use of) terms such as 'seeing' as if they were 'achievements or triumphs' (Sibley 1955, 471). We do so precisely when we describe a scene for an agent to act in – e.g. a field in which there are cows – such that we are then able to claim or assert whether the agent 'achieves' the seeing of the cows (or, for some reason to be investigated, sees the cows mistakenly as small, fat horses). Sibley then attempts – to help us out of this correlation of seeing with achievement – to find another sense of 'seeing':

> The sense of 'see' I have in mind is that in which, to look at or scrutinise an object for a given length of time, one must throughout that length of time be seeing it. I cannot scrutinise something unless I can, at the time, actually see it (though of course I may see it without scrutinising it). If for any period of time I am unable to see the object in question, for that length of time I cannot scrutinise it or examine. If 'he is now watching it' is true, 'he can now see it'

must also be true. So here is at least one use of 'see' which is clearly different from the achievement or spotting use. (Sibley 1955, 472)

The reference to a period of time – to some duration – is important. Part of what is going on in the case of understanding 'seeing' as an achievement or triumph is a kind of stop-start image of cognition at work, as if cognition was constantly making one photograph after another (such that the film could then be appealed to justify what inferences one drew from 'what' one photographed). Sibley's point helps us see that if we think of it as a process occurring over time, it may be possible to resist this urge to impose an achievement meaning on the term 'seeing'. One should add here that the appeal to duration does not itself to do the trick, for someone can – as Sibley points out – still treat duration as a specific time period, e.g. seeing from 2pm to 2.10pm (Sibley 1955, 472), at which point the image of triumph or achievement returns (the image there might be one of me being able to keep my eye on some fast-moving object for 10 minutes – an achievement or triumph of a kind). One begins to see the need for something open, or open-ended – which is enabled only precisely when we do not presuppose things or ourselves to be in a certain way.

Sibley, however, sometimes speaks of 'seeing' (in the above sense absent of an achievement reading) as passive or doing something without effort or difficulty. His contrast case is precisely one of keeping an eye on something, so he says: 'Keeping it in sight is a task or activity, something one does, sets oneself to do, engages in. It can be done skilfully, carefully, resolutely, doggedly, with all one's strength and with great effort and expenditure of energy. None of this is true of simply seeing something' (Sibley 1955, 474). Here we have to disagree with Sibley. The very point is to understand seeing – in the sense of attentive silence – as something active, but not necessarily active in the sense of being an activity familiar to us, and one we know when it is complete.

The fact that there is a sense of seeing that is not an achievement does not mean it is not active, nor indeed something that requires some effort, and thus something that can be difficult – indeed very difficult – and that can certainly be done skilfully, carefully, resolutely, and even doggedly. I can will myself to attend silently, and I can continue to do so, resisting the pull (if I experience it as a pull) of appreciating it or claiming / asserting something about it. One can have this experience of willing oneself to attend silently (as silently as one can) perhaps most

clearly in the case of focusing in great concentration on an abstract problem, i.e. one does not so much move through possible ways of solving it, or possible ways of thinking what the problem demands (what kind of problem it is), and yet be wrestling with it (as if turning it over in one's mind). Similarly, I can close my eyes and run my hand over a fabric, and let it experience the fabric, while trying hard to not hurry to impose a certain way of appreciating it or judging it to be in a certain manner.

C. Processes without End(s)

Earlier, in the last Chapter of the preceding Part (see Section 23.E.), we mentioned that one of the interesting issues that arose in a recent exchange (between Green 2010 and Stout 2010) in the context of inter-personal perception was the extent to which one might have to think – if one wants to describe persons as perceiving the emotions of others – of that ability to perceive others' emotions as processual, rather than in some stop-start (photographic) manner. The point was made by Stout (2010), who invoked the difference between seeing a signature and seeing someone signing: on the signature view, which he attributes to Green (2010), we recognise someone as having a certain emotion (i.e. that they angry); on the seeing someone signing view, we 'see their anger as it plays out on their face' – or, put in general terms: 'The possibility to consider here is that one can witness someone's expressing their emotional state in the development of their facial pattern even though one may not be able to witness their state just by examining that facial pattern' (Stout 2010, 35).

Stout is concerned in the paper with various issues not directly relevant for present purposes, but the above reference to the difference between perceiving someone as angry in some moment, and perceiving their anger as it plays out on their face, allows us to make the following observation. Stout is surely right in helping us see how differently the experience of a process is from a 'product' (as he calls it). But the crucial issue here is whether we think of that process as precisely one that must culminate in – or involve the 'playing out' or enactment of – some particular emotion, e.g. anger. If one sees 'process' as but another pattern – just one elongated in time – and one that either enacts some pattern or culminates in a particular product, one then performs the very same reduction that a singular time-frame product does.

To take process seriously must mean precisely not perceiving the other's face as 'enacting' some emotion, or as leading up to (as if building up to) some final snapshot of anger. Certainly, one can have the experience that someone's face is

building up to anger, or even playing out some pattern one can associate with anger; but if one wants to take process seriously then one must do so in an open way. Doing so in an open way involves making room both for the possibility of attending to their face silently, and to an ongoing relation with them where one appreciates how they might be feeling (again, at some point during this relation one might assert or claim, in a justifiable or systematisable way, that the face is angry, but it is vital to see that this is not the only thing one can do).

Stout's reminder of the importance of process is a good one, but the way he uses it suggests he thinks of us as stocked with a catalogue of processes of certain emotions, such that what we see is determined by that stock (even if only by defeasible default): we either see, by default, processes culminating in certain kinds of emotions, or we see the enactment of certain kinds of emotion-processes. To use the notion of process in that way would be to miss the phenomenon of attentive silence.

D. Non-Representational

The way we are describing attentive silence can be thought to be related to the debate concerning the extent to which, if at all, experiences represent, i.e. whether experiences necessarily (or sometimes) have representational content. This debate can overlap with the conceptual / non-conceptual debate,[249] for the issue of representation can be read as concerning whether 'conscious experience presents to us a range of things and properties that can outstrip the range of things and properties available to us in thought' (where 'in thought' is read as enabled by and requiring the exercise of conceptual faculties; the quote is from Luntley 2010, 217). Here, what is 'available to us in thought' can be read as a matter of being enabled by and requiring the exercise of conceptual faculties, but it can also be divorced from that debate – as in fact Luntley argues it ought to be – by being a claim about content available to thought in some more basic way. Luntley leans heavily on psychological research on the experience of music, the basic point being that 'our conscious experience of music can provide us with awareness of various features of music without our having the capacity to think about those features' (Luntley 2010, 218).

The way Luntley uses 'awareness without content' here can be likened to 'attentive silence', and thus this may be a resource future work on attentive silence

---

[249] That debate was addressed, to some extent, in Chapter 12 (which focused on McDowell's view).

could build on. The one proviso to make about this work concerns what is understood, in the quote just transcribed, as 'features', i.e. the key question is whether the theorist is assuming (for the purposes of describing the person's awareness) that the music has these and only these features that the person can be aware of (without this being available to thought – and thus without being represented in that way). If the theorist is assuming that there are such-and-such 'features' to be recognised and no others (as Luntley seems to) then they will describe the cognition of the agent in more rigid terms than we want to: the point is that what attentive silence enables (when it enables appreciation) is the creation of what may, from then on, count as a feature.

We can describe a person experiencing some piece of music that we understand to have certain features and assert that the person may experience those features without them being available to thought – this is indeed a valuable observation, but not quite the point of attentive silence. Instead, part of the point of attentive silence is to tell us that we need to avoid describing, in advance of the agent's engagement with the environment, the environment (in this case the piece of music) to be in a certain way (to have certain features that await the agent's recognition of them), and this is precisely because as soon as we confine our description of an agent's cognition to recognition we make it out to be much less active and creative than it can be (we fail to see how the person could be consciously aware of (in Luntley's terms), or attending to (in our terms), the piece of music in ways that we have never identified before, and thus not in terms of what we might pre-describe as 'the features' of that piece of music.[250]

E. Un-Situated

Let us also reiterate that the notion of attentive silence can help us loosen what we might – especially when we want to ward off views like the conceptual content view, or representationalist views – otherwise treat as too closely tied together, i.e. especially action and perception. In this context, one of the most prominent

---

[250] The same points that are made here by reference to Luntley's views may also apply to views that base perceptual knowledge on an ability to discriminate, e.g. Millar 2000, building on Goldman 1976 – the point about these contributions is that in order to avoid deliberation and inference-making, these views end up giving an account that is too mechanical: they end up offering an account of perception that restricts it to recognising the same differences, and this pictures our perceptual capacities as much too passive and rigid. The point being made here is that our cognition needs to be understood as capable of creating what might come to be treated, by us or by others, as a relevant difference.

approaches in the literature is that of the theory of situated cognition (see Robbins and Aydede 2009), as well as various treatments of perception as a kind of action (see Noë 2004). As Andy Clark (2006) points out, Noë's view requires having 'implicit knowledge of…rules or regularities relating sensory inputs to movement' (Clark 2006, 2). Appeals here are often made to 'motor intentionality', often by reference to Dreyfus' work (e.g. see Kelly 2000), where the focus is on 'skilled motor behaviours that take place naturally in the context of environmental cues' (Kelly 2000, 165), and more recently by reference to dynamic systems theory (e.g. see Schöner and Reimann 2009).

These moves are often made in order to avoid over-intellectualising the mind, i.e. in the sense of avoiding seeing it as involved in computational or deliberative tasks (as if the mind was always behind, directing and manipulating a puppet of a body). These moves are important, as they help us avoid dualisms that have pernicious effects, but the problem, from our perspective, in tying perception and action together too tightly as a response to over-intellectualisation, is that the resultant views can offer just another version of cognitive passivity and rigidity, e.g. in situated theories, we are precisely situated in an environment that is always and already meaningful for us given the specific ways in which we allegedly learn to find things meaningful (especially as experts in specific domains).

Certainly, we need to make room for the ways in which persons need not engage in computational or deliberative tasks (developing, for instance, a kind of bodily memory), but the problem with the situated view (and others that tie perception tightly to action – e.g. 'if it itches, scratch!': see Hall 2008) is that it leaves open the possibility of asserting that in acting in ways that are non-computational and non-deliberative we have internalised the rules (e.g. rules for how things ought to be used or responded to, as determined by 'the situation' we find ourselves in) or asserting that we have, as Searle says, acquired patterned ways of doing things that happen to be in accordance with the rules. The effect of this is to describe us as much more passive and rigid than we are. The point of attentive silence is to help us avoid that kind of response to over-intellectualisation, and so it is vital that we do not characterise attentive silence as a matter of immediate perception of things as

meaningful, or as perception of things in such a way that we immediately experience a call to do something.[251]

F. The Partnership with Appreciation

Finally, one final word ought to be devoted to stating clearly, at the risk of repetition, that cognition from the second-person perspective is a partnership between appreciation and attentive silence. As has been observed throughout, both appreciation and attentive silence are tentative, hesitant and, therefore, also risky engagements with the environment and others. As a result, they are also both flexible and inherently open to revision. They are also non-committal and non-assertive, though this has here been conceptualised as a matter of degree, so that although appreciation is minimally committal and assertive, it is still more committal and assertive than attentive silence.

Despite these similarities, there is also an important difference: attentive silence can occur in the absence of appreciation, whereas it is difficult to imagine appreciation operating without being preceded by attentive silence (at least in the majority of cases). Both before and during the course of interaction (if indeed we are interacting) we attend silently to someone – we open ourselves up to experiencing them in a way we have not previously experienced them or any another person, including possibly experiencing them (or what they do) as appropriate or inappropriate. When we attend silently to another we are being active, and we are engaging with them richly (for we can be surprised by how we experience them), but we are not being so active as to give that experience any particular form. We should also add that we can attend silently to someone and yet accompany this with certain (normative) anticipations or expectations of them – there is no reason, then, why attentive silence cannot operate side by side with the treatment of certain normative resources as relevant to that interaction.

This period of attentive silence – accompanied or not by contingently useful normative resources – prepares us for, or enables us, to then fashion that experience into the form of either the very tentative and hesitant experience of appreciation, or the much more assertive and committal (and also status-anxious, or status-seeking)

---

[251] It may be that there are ways of making the intuitions in situationism work without defining attentive silence out of existence or making it marginal. To consider this would require delving into detail into the literature on situated cognition, and this, regrettably, is a task that lies outside the scope of this thesis.

form of justification and / or systematisation. We could say that attentive silence is actively passive, whereas appreciation is passively active. They work together well in interaction understood from the second-person perspective, for they enable a person precisely to not only tolerate, but also to make positive use of, the irregular, the ambiguous and the unpredictable in any interaction. In that sense, they also enable a person to relate to, to be more responsive to, the difference of others – that otherness of others that was discussed in Chapter 23 in Part V.

The model of the partnership is general, but it is also applicable to normative experience. Persons can experience each other's expressions and actions as perhaps appropriate or inappropriate – this is the act of appreciation. Prior to that, they can be attending silently to each other, experiencing each other's potential appropriateness or inappropriateness richly, and yet without giving that experience any particular form. They thereby enable themselves to create, for each other, what might be considered to be relevant in future claims (justifiable and / or systematisable) concerning something someone does as appropriate or inappropriate. One could say that what attentive silence and appreciation working together indicate is the workings of the emergence of new relevancies, including precisely relevancies for experiences as to when something is appropriate or inappropriate and what action ought to follow. In that sense, the partnership between appreciation and attentive silence is important not only for participation in social life, understood from the second-person perspective, in a general way, but also in its normative dimension.

Conclusion

In concluding this thesis, let us return to our epigraph, and thus to Khushim. From the second-person perspective, Khushim's problem was not that he did not have adequate knowledge of social norms, including knowing when they are applicable. Instead, Khushim's problem was that he was incapable of experiencing tentativeness and hesitation in interaction with others. He was intolerant of any ambiguity and unpredictability in his relations with others. He was incapable, in short, of attending silently to others and appreciating how it might be appropriate or inappropriate to relate with them. He was, it seems, incapable of learning from others how to relate with them in the course of interacting with them. He rushed in, following one rule after another, as if there were no persons there at all – as if others were just signs for the application of rules he had just acquired, or characters in a scripted play, or toys to be moved around in some configuration.

Arguably, Khushim is not the only victim of the story. The others in the folktale are also victims – not the victims of Khushim's insults, but of the description of their actions: for is it so obvious that they would react to Khushim by beating him soundly? Could we not imagine someone in one of those groups engaging with Khushim, saying 'slow down', 'hold on', giving him an opportunity to relate, at least learn to relate with others? Would Khushim really be destined to inevitable beatings, to be treated as a violator, his difference a negative one and one marked by incapacity to comply with the rules allegedly held dear by the members of the community? In other words, are not both Khushim and those others who beat him caricatured in the tale? And if they are, why are they so caricatured – is it really to show us the importance of social norms, or is it to show us their inadequacy (both for Khushim and for understanding those relating with him)?

Perhaps, in the repetition of the motif (of Khushim not succeeding in relating to others) the folktale shows us what it could not tell us: that there must be something more to participating successfully in social life than acquiring knowledge of, and thereafter applying social norms. It could not tell us this straight out because we could dismiss any one example as a simple misapplication of the norm. We could attribute it, then, to the incomplete normative knowledge that Khushim has. But this reading – because of the repetition of the motif – does not sit so comfortably. Instead, it points precisely to something more – some other missing ability or abilities to relate to

others. This thesis has sought to show, first, the inadequacy of an approach to our understanding of how we manage, on the whole successfully, to get along with others based on our acquisition and internalisation of rules; and second, the need for something to supplement the contingent usefulness of normative resources, namely, the partnership between appreciation and attentive silence.

We need, then, to be very careful in how we approach accounting for the normative dimension of social life. We cannot account for it in such a way that will exclude the irregularity, ambiguity, unpredictability and uncertainty that is such an important part of social life. We also cannot account for it in such a way that it would exclude the task of describing our abilities to participate in social life so conceived, i.e. abilities that are more hesitant and tentative in character. Perhaps, ultimately, there is a plea made in this thesis for recognition, and careful description, of a range of experiences, attitudes and abilities that are less assertive and less committal; more a matter of doubt than confidence; more gentle than powerful. This is not to idealise – instead, it is to see value in that which we all too often dismiss as an obstacle, something to be overcome. It is precisely to try to end not with a bang, but with a whisper.

Bibliography

Abrahamsen, A, and W, Bechtel. (2006), 'Phenomena and Mechanisms: Putting the Symbolic, Connectionist, and Dynamical Systems Debate in Broader Perspective', in R. Stainton (ed.), *Contemporary Debates in Cognitive Science*, Oxford: Basil Blackwell, 159-185

Albritton, R. (1959) 'On Wittgenstein's Use of the Term "Criterion"', *The Journal of Philosophy*, 56: 22, 845-857

Andrews, C. (2003), 'Constraint and Convention: The Formalism of the Oulipo', *Neophilologus*, 87: 2, 223-232

Arvidson, S. (2003), 'Moral Attention in Encountering You: Gurwitsch and Buber', *Husserl Studies*, 19: 71-91

Audi, R. (2010), 'Moral Perception and Moral Knowledge', *Aristotelian Society Supplementary Volume*, 84: 1, 79-97

Austin, J. (1979 [1956]), 'A Plea for Excuses', in *Philosophical Papers*, ed. by J. Urmson and M. Warnock, Oxford: Oxford University Press, 175-204

Austin, J.L. (1962), *Sense and Sensibilia*, Oxford: Oxford University Press

Baerveldt, C and P. Voestermans. (2005), 'Culture, Emotion and the Normative Structure of Reality', *Theory & Psychology*, 15: 4, 449-473

Baerveldt, C and T. Verheggen. (2007), 'We Don't Share! The Social Representation Approach, Enactivism and the Ground for an Intrinsically Social Psychology', *Culture & Psychology*, 13, 5-27

Bakhtin, M. (1984 [1963]), *Problems of Dostoyevsky's Poetics*, ed. and trans. by C. Emerson, Minneapolis: University of Minnesota Press

Bekoff, M. (2001), 'Social Play Behaviour: Cooperation, Fairness, Trust and the Evolution of Morality', *Journal of Consciousness Studies*, 8, 81-90

Bennett, J. (1989), *Rationality: An Essay toward Analysis*, Indianapolis: Hackett

Bernstein, R. (2010), *The Pragmatic Turn*, Cambridge: Polity

Bicchieri, C. (2006), *The Grammar of Society*, Cambridge: Cambridge University Press

Bickhard, M. H. (2009), 'The Interactivist Model', *Synthese*, 166: 3, 547-591

Bigelow, A. E. and P. Rochat. (2006), 'Two-Month-Old Infants' Sensitivity to Social Contingency in Mother-Infant and Stranger-Infant Interaction', *Infancy*, 9: 3, 313-325

Bilgrami, A. (2005), 'Rules', in Honderich (ed.), 824-825

Black, M. (1962), *Models and Metaphors: Studies in Language and Philosophy*, Ithaca: Cornell University Press

Blum, L. (1994), *Moral Perception and Particularity*, Cambridge: Cambridge University Press

Boden, M. (2006), 'Of Islands and Interactions', *Journal of Consciousness Studies*, 13: 5, 53-63

Boghossian, P. (1989), 'The Rule-Following Considerations', *Mind*, 98: 392, 507-549

Boghossian, P. (2003), 'The Normativity of Content', *Philosophical Issues*, 13: 1, 31-45

Boghossian, P. (2008a), 'Epistemic Rules', *The Journal of Philosophy*, 105: 9, 472-500

Boghossian, P. (2008b), *Content and Justification: Philosophical Papers*, Oxford: Oxford University Press

Bohman, J. (2000), 'The Importance of the Second Person: Interpretation, Practical Knowledge and Normative Attitudes', in H.H. Kogler and K.R. Stueber (eds.), *Empathy and Agency: The Problem of Understanding in the Human Sciences*, Philadelphia: Westview Press, 222-242

Bourdieu, P. (1990), *The Logic of Practice*, trans. by R. Nice, Cambridge: Polity Press

Brady, M. (2010), 'Virtue, Emotion and Attention', *Metaphilosophy*, 41: 1, 115-131

Brandom, R. (1994), *Making it Explicit: Reasoning, Representing and Discursive Commitment*. Cambridge, MA: Harvard University Press

Broom, L. and P. Selznick. (1963), *Sociology: A Text with Adapted Readings*, 3rd ed., New York: Harper & Row

Broome, J. (2005), 'Does Rationality Give Us Reasons?', *Philosophical Issues*, 15, 321-327

Broome, J. (2007), 'Wide or Narrow Scope?', *Mind*, 116: 462, 359-370

Broome, J. (Unpublished), 'Ought', Draft made available as part of an end-of-semester Seminar Workshop at the Department of Philosophy, University of Geneva, 2009

Buber, M. (1970), *I and Thou*, trans. by W. Kaufmann, New York: Simon and Schuster

Burge, T. (2010), *Origins of Objectivity*, Oxford: Oxford University Press

Burrows, J. (2007), *A History of Histories: Epics, Chronicles, Romances and Inquiries from Herodotus and Thucydides to the Twentieth Century*, London: Allen Lane

Cassam, Q. (2007), *The Possibility of Knowledge*, Oxford: Clarendon Press

Chappell, T. (2008), 'Moral Perception', *Philosophy*, 83: 4, 421-437

Cheal, D. (1980), 'Rule-Governed Behaviour', *Philosophy of the Social Sciences*, 10: 1, 39-49

Chomsky, N. (1980), 'Rules and Representations', *Behavioural and Brain Sciences*, 3, 1-15

Chomsky, N. (1996 [1988]) 'Language and Problems of Knowledge', in A.P. Martinich (ed.), *The Philosophy of Language*, 3rd ed., Oxford: Oxford University Press, 558-578

Church, J. (2010), 'Seeing Reasons', *Philosophy and Phenomenological Research*, 80: 3, 638-670

Cialdini, R. B., R.R. Reno, and C.A. Kallgren. (1990). 'A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places', *Journal of Personality and Social Psychology*, 58, 1015-1026

Clark, A. (2006), 'Vision as Dance? Three Challenges for the Sensorimotor Contingency Theory', *Psyche*, 12: 1, 1-10

Cohen, D. (2006), *The Development of Play*, 3rd ed., London: Routledge

Cole, J. (1998), *About Face*, Cambridge, Mass.: MIT Press

Collins, H. (2010), *Tacit and Explicit Knowledge*, Chicago: Chicago University Press

Colombetti, G. (2007), 'Enactive Appraisal', *Phenomenology and the Cognitive Sciences*, 5, 527-546

Coyne, M. (1981), 'Beyond Rules: Mapping the Normative', *American Philosophical Quarterly*, 18: 4, 331-337

Cullison, A. (2010), 'Moral Perception', *European Journal of Philosophy*, 18: 2, 159-175

D'Arms, J and D. Jacobson. (2006), 'Sensibility Theory and Projectivism', in D. Copp (ed.), *Oxford Handbook of Ethical Theory*, New York: Oxford University Press, 186-219

Dalton, B. (2004), 'Creativity, Habit, and the Social Products of Creative Action: Revising Joas, Incorporating Bourdieu', *Sociological Theory*, 22: 4, 603–622

Dancy, J. (2004), *Ethics Without Principles*, Oxford: Oxford University Press

Dancy, J. (2010), 'Moral Perception', *Aristotelian Society Supplementary Volume*, 84: 1, 99-117

Darwall, S. (2006), *The Second-Person Standpoint: Morality, Respect, and Accountability*, Cambridge, Mass.: Harvard University Press

Darwall, S. (2010), 'Précis: *The Second-Person Standpoint*', *Philosophy and Phenomenological Research*, 81: 1, 216-228

Davidson, D. (1992), 'The Second Person', in P.A. French, T.E. Uehling and H.K. Wettstein (eds.), *Midwest Studies in Philosophy, Volume XVII: The Wittgenstein Legacy*, Notre Dame: University of Notre Dame Press, 255-267

De Certeau, M. (1984), *The Practice of Everyday Life*, Berkeley: University of California Press

De Jaegher, H and E. Di Paolo. (2007), 'Participatory Sense-Making: An Enactive Approach to Social Cognition, *Phenomenology and the Cognitive Sciences*, 6, 485-507

De Jaegher, H and T. Froese (2009). 'On the Role of Social Interaction in Individual Agency', *Adaptive Behavior*, 17: 5, 444-460

De Jaegher, H, E. Di Paolo, and S. Gallagher. (2010), 'Can Social Interaction Constitute Social Cognition?', *Trends in Cognitive Sciences*, 14: 10, 441-447

De Jaegher, H. (2006), *Social Interaction Rhythm and Participatory Sense-Making: An Embodied, Interactional Approach to Social Understanding, with Some Implications for Autism*, DPhil, Sussex: University of Sussex

de Sousa, R. (1987), *The Rationality of Emotion*, Cambridge, Mass.: MIT Press

de Sousa, R. (2002), 'Emotional Truth', *Proceedings of the Aristotelian Society*, 76: 1, 247-263

de Vignemont, F. (2008), 'Frames of Reference in Social Cognition', *Quarterly Journal of Experimental Psychology*, 61: 1, 90-100

Del Mar, M. (2010), 'Normativism, Anti-Normativism and Humanist Pragmatism', *Human Studies*, 33: 2-3, 305-323

Del Mar, M. (2011a), 'Marmor's *Social Conventions*: The Limits of Practical Reason', *Philosophy of the Social Sciences*, 41: 3, 420-445

Del Mar, M. (2011b), 'Concerted Practices and the Presence of Obligations: Joint Action in Competition Law and Social Philosophy', *Law & Philosophy*, 30: 1, 105-140

Del Mar, M. (2011c), 'What Does History Matter to Legal Epistemology?', 5 *Journal of the Philosophy of History* 383-405

Dent, N. (2005), 'Normative', in Honderich (ed.), 662

Deonna, J. (2006), 'Emotion, Perception and Perspective', *Dialectica*, 60: 1, 29-46

Deonna, J and F. Teroni. (2012), *The Emotions: A Philosophical Introduction*, London: Routledge

Dequech, D. (2009), 'Institutions, Social Norms, and Decision-Theoretic Norms', *Journal of Economic Behavior and Organization*, 72: 1, 70-78

DesAutels, P. (1996), 'Gestalt Shifts in Moral Perception', in L. May, M. Friedman and A. Clark (eds.), *Mind and Morals: Essays on Cognitive Science and Ethics*, Cambridge, Mass.: MIT Press, 129-143

Devries, R. (1998), 'Games with Rules', in D.P. Fromberg and D. Bergen (eds.), *Play from Birth to Twelve and Beyond: Contexts, Perspectives, and Meanings*, New York: Garland Publishing, 409-415

Dharamsi, K (2011), 'Re-enacting in the Second Person', *Journal of the Philosophy of History*, 5, 163-178

DiCicco-Bloom, B and David R. Gibson. (2010), 'More than a Game: Sociological Theory from the Theories of Games', *Sociological Theory*, 28: 3, 247–271

Döring, S. (2007), 'Seeing What to Do: Affective Perception and Rational Motivation', *Dialectica*, 61: 3, 363-394

Downing, G, D. Bürgin, C. Reck, and U. Ziegenhain. (2008), 'Interfaces Between Intersubjectivity and Attachment: Three Perspectives on a Mother-Infant Inpatient Case', *Infant Mental Health Journal*, 29: 3, 278-295

Dretske, F. (1973), 'Perception and Other Minds', *Nous*, 7: 34–4

Dreyfus, H and S. Dreyfus. (1991), 'Toward a Phenomenology of Ethical Expertise', *Human Studies*, 14, 229-250

Dreyfus, H. (1972), *What Computers Can't Do*, Cambridge, Mass.: MIT Press

Dreyfus, H. (2006), 'Overcoming the Myth of the Mental', *Topoi*, 25: 1-2, 43-49

Dreyfus, H. (2007a), 'The Return of the Myth of the Mental', *Inquiry*, 50: 4, 352-365

Dreyfus, H. (2007b), 'Response to McDowell', *Inquiry*, 50: 4, 371-377

Ellsworth, P.C. and K.R. Scherer. (2003), 'Appraisal Processes in Emotions', in R.J. Davidson, H. Goldsmith and K.R. Scherer (eds.), *Handbook of Affective Sciences*, New York / Oxford: Oxford University Press, 572-595

Elster, J. (1989), *The Cement of Society: A Study of Social Order*, Cambridge: Cambridge University Press

Elster, J. (1994), 'Rationality, Emotions and Social Norms', *Synthese*, 98: 1, 21-49

Elster, J. (2007), *Explaining Social Behaviour: More Nuts and Bolts for the Social Sciences*, Cambridge: Cambridge University Press

Enoch, D. (2006), 'Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action', *Philosophical Review*, 115: 2, 169-198

Evans, G. (1982), *Varieties of Reference*, J. McDowell (ed.), Oxford: Oxford University Press

Favereau, I. (2008), 'The Unconventional, but Conventionalist, Legacy of Lewis's "Convention"', *Topoi*, 27: 1-2, 115-126

Finlay, S. (2006), 'The Reasons that Matter', *Australasian Journal of Philosophy*, 84, 1-20

Finlay, S. (2010), 'Recent Work on Normativity', *Analysis*, 70: 2, 331-346

Fodor, J. (1983), *The Modularity of Mind*, Cambridge, Mass.: MIT Press

Fogel, A. (1993), *Developing through Relationships*, Chicago: University of Chicago Press

Freeman, J. (2011), 'LOL and the O.E.D.', *International Herald Tribune*, Monday April 18 2011, 9

Fuchs, T and H. De Jaegher. (2009), 'Enactive Intersubjectivity: Participatory Sense-Making and Mutual Incorporation, *Phenomenology and the Cognitive Sciences*, 8: 4, 465-486

Gallagher, S and D. Hutto. (2008), 'Understanding Others Through Primary Interaction and Narrative Practice', in J. Zlatev et al (eds.), *The Shared Mind: Perspectives on Intersubjectivity*, Amsterdam: John Benjamins Publishing Company, 17-38

Gallagher, S and D. Zahavi. (2008), *The Phenomenological Mind*, London: Routledge

Gallagher, S. (2008a), 'Direct Perception in the Intersubjective Context', *Consciousness and Cognition,* 17: 2, 535–543.

Gallagher, S. (2008b), 'Intersubjectivity in Perception', *Continental Philosophy Review*, 41, 163-178

Gallagher, S. (2008c), 'Inference or Interaction: Social Cognition Without Precursors', *Philosophical Explorations*, 11: 3, 163-74

Gallagher, S. (2009), 'Two Problems of Intersubjectivity', *Journal of Consciousness Studies*, 16: 6-8, 289-308

Ganz, J.S. (1971), *Rules: A Systematic Study*, The Hague: Mouton

Garfinkel, H. (1967), *Studies in Ethnomethodology*, New Jersey: Prentice

Geertz, C. (1973), *The Interpretation of Cultures: Selected Essays*, New York: Basic Books

Geertz, C. (1983), *Local Knowledge*, New York: Basic Books

Gergely, G. (2001), 'The Obscure Object of Desire: "Nearly, but Clearly Not, Like Me": Contingency Preference in Normal Children versus with Autism', *Bulletin of the Menningner Clinic*, 65: 3, 411-426

Gibbard, A. (1990), *A Theory of Normative Judgement*, Cambridge, Mass.: Harvard University Press

Gibbard, A. (2003), *Thinking How to Live*, Cambridge, Mass.: Harvard University Press

Gibbs, J. (1965), 'Norms: The Problem of Definition and Classification', *The American Journal of Sociology*, 70: 5, 586-594

Giddens, A. (1984), *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press

Gilbert, M. (1989), *On Social Facts*, Princeton: Princeton University Press

Gilbert, M. (1990), 'Walking Together: A Paradigmatic Social Phenomenon', *Midwest Studies in Philosophy*, 15: 1, 1-14

Gilbert, M. (2008), 'Social Conventions Revisited', *Topoi*, 27: 1-2, 5-16

Goffman, E. (1983), 'The Interactive Order', *American Sociological Review*, 48: 1, 1-17

Goffman, E. (1961), *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*, New York: Doubleday

Goffman, E. (1963), *Behaviour in Public Places: Notes on the Social Organisation of Gatherings*, New York: The Free Press

Goffman, E. (1967), *Interaction Ritual: Essays on Face-to-Face Behaviour*, New York: Anchor Books

Goffman, E. (1970), *Strategic Interaction*, Philadelphia: University of Pennsylvania Press

Goffman, E. (1971), *Relations in Public: Microstudies of the Public Order*, New York: Basic Books

Goldie, P. (2007), 'Seeing What is the Kind of Thing to Do: Perception and Emotion in Morality', *Dialectica*, 347-361

Goldman, A. (1976), 'Discrimination and Perceptual Knowledge', *The Journal of Philosophy*, 73: 20, 771-791

Green, M. (2007), *Self-Expression*, Oxford: Oxford University Press

Green, M. (2010), 'Perceiving Emotions', *Aristotelian Society Supplementary Volume*, 84: 1, 45-61

Grossen, M. (2010), 'Interaction Analysis and Psychology: A Dialogical Perspective', *Integrative Psychological and Behavioural Science*, 44, 1-22

Grimshaw, A. (1980), 'Social Interactional and Social Linguistic Rules', *Social Forces*, 58: 3, 789-810

Habermas, J. (1996), *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. by W. Rehg, Cambridge, Polity Press

Hahn, U. and N. Chater. (1998), 'Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable?', *Cognition*, 65, 197-230

Hall, R. (2008), 'If It Itches, Scratch!', *Australasian Journal of Philosophy*, 86: 4, 525-535

Hanna, R. (2006), 'Rationality and the Ethics of Logic', *Journal of Philosophy*, 103, 67-100

Haugeland, J. (1998), *Having Thought: Essays in the Metaphysics of Mind*, Cambridge, Mass: Harvard University Press

Hayek, F.A. (1967), *Studies in Philosophy, Politics and Economics*, London: Routledge & Kegan Paul

Heal, J. (2007), 'Back to the Rough Ground! Wittgensteinian Reflections on Rationality and Reason', *Ratio*, 20, 403-421

Henderson, D. (2002), 'Norms, Normative Principles, and Explanation', *Philosophy of Social Science*, 32, 329-364

Herschbach, M. (2008), 'Folk Psychological and Phenomenological Accounts of Social Perception', *Philosophical Explorations*, 11: 3, 223-235

Hershfield, J. (2005), 'Rule Following and the Background', *Linguistics and Philosophy*, 28: 3, 269-280

Hobson, P. (1995), *Autism and the Development of Mind*, East Sussex: Lawrence Erlbaum Associates

Hobson, P. (2002), *The Cradle of Thought: Exploring the Origins of Thinking*, London: Macmillan

Honderich, T (ed.). (2005), *The Oxford Companion to Philosophy*, 2nd ed., Oxford: Oxford University Press

Horgan, T and M. Timmons. (2009), 'What Does the Frame Problem Tell us About Moral Normativity?', *Ethical Theory and Moral Practice*, 12: 1, 25-51

Horwich, P. (2010), *Truth, Meaning, Reality*, Oxford: Oxford University Press

Humberstone, I.L. (1992), 'Directions of Fit', *Mind*, 101: 401, 59-83

Hutto, D. (2004), 'The Limits of Spectatorial Folk Psychology', *Mind & Language*, 19: 5, 548-573

Hutto, D. (2008), *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, Mass.: MIT Press

Ingold, T. (2007), *Lines: A Brief History*, London: Routledge

Jacobson, D. (2005), 'Seeing by Feeling: Virtues, Skills and Moral Perception', *Ethical Theory and Moral Practice*, 8, 387-409

Jacovides, M. (2010), 'Do Experiences Represent?', *Inquiry*, 53: 1, 87-103

Jerolmack, C. (2009), 'Humans, Animals, and Play: Theorizing Interaction When Intersubjectivity is Problematic', *Sociological Theory*, 27: 4, 371–389

Joas, H. (1996), *The Creativity of Action*, Cambridge: Polity Press

Johnston, M. (2001), 'The Authority of Affect', *Philosophy and Phenomenological Research*, 63: 1, 181-214

Kallgren, C. A., R.R. Reno, and R.B. Cialdini. (2000), 'A Focus Theory of Normative Conduct: When Norms Do and Do Not Affect Behaviour', *Personality and Social Psychology Bulletin*, 26, 1002-101

Kelly, S.D. (2000), 'Grasping at Straws: Motor Intentionality and the Cognitive Science of Skilful Action', in M. Wrathall and J. Malpas (eds.), *Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus*, Vol. II, Cambridge, Mass.: MIT Press, 161-177

Kelso, S. (1995), *Dynamic Patterns: The Self-Organisation of Brain and Behaviour*, Cambridge, Mass.: MIT Press

Kew, F. (1986), 'Playing the Game: An Ethnomethodological Perspective', *International Review for the Sociology of Sport*, 21: 4, 305-322

Kew, F. (1987), 'Contested Rules: An Explanation of How Games Change', *International Review for the Sociology of Sport*, 22: 2, 125-135

Kew, F. (1992), 'Game-Rules and Social Theory', *International Review for the Sociology of Sport*, 27: 4, 293-307

Kinda, Z. (1999), *Social Cognition: Making Sense of People*, Cambridge, Mass.: MIT Press

Klaassen, P, E. Rietveld, and J. Topal. (2010), 'Inviting Complementary Perspectives on Situated Normativity in Everyday Life', *Phenomenology and the Cognitive Sciences*, 9: 1, 53-73

Kohlberg, L. (1986), *The Philosophy of Moral Development*, San Francisco: Harper and Row

Kolodny, N. (2005), 'Why be Rational?' *Mind*, 114, 509-563

Korsgaard, C. (1997), 'The Normativity of Instrumental Reason', in B. Cullity and B.N. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press, 215-254

Korsgaard, C. (2008), *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford: Oxford University Press

Kripke, S. (1982), *Wittgenstein on Rules and Private Language*, Cambridge, Mass: Harvard University Press

Latsis, J. (2005), 'Is There Redemption for Conventions?', *Cambridge Journal of Economics*, 29, 709-727

Lavin, D. (2004), 'Practical Reason and the Possibility of Error', *Ethics*, 114, 424-57

Lecomte, A. and M. Quatrini. (2011), 'Figures of Dialogue: A View from Ludics', *Synthese*, 183: Supplement 1, 59-85

Lewis, D. (1986), *Convention*, Oxford: Blackwell

Lewis, J.D. and J.L. Elman. (2001), 'A Connectionist Investigation of Linguistic Arguments from the Poverty of Stimulus: Learning the Unlearnable', *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 552-557

Linell, P. (2003), 'Dialogical Tensions: On Rommetveitian Themes of Minds, Meanings, Monologues and Languages', *Mind, Culture, and Activity* 10: 3, 219-229

Linell, P. (2009), *Rethinking Language, Mind and World Dialogically: Interactional and Contextual Theories of Human Sense-Making*, Charlotte: Information Age Publishing

Louise, J. (2009), 'Correct Responses and the Priority of the Normative', *Ethical Theory and Moral Practice*, 12, 345-364

Luntley, M. (2010), 'Expectations Without Content', *Mind & Language*, 25: 2, 217-236

Manning, P. (1992), *Erving Goffman and Modern Sociology*, Cambridge: Polity

Marková, I. (2003), *Dialogicality and Social Representations: The Dynamics of Mind*, Cambridge: Cambridge University Press

Marmor, A. (2001) *Positive Law and Objective Values*, Oxford: Oxford University Press

Marmor, A. (2006), 'How Law is Like Chess', *Legal Theory*, 12, 347-371

Marmor, A. (2009), *Social Conventions: From Language to Law*, Princeton: Princeton University Press

Mauss, M. (1979), 'Body Techniques', in *Sociology and Psychology: Essays by Marcel Mauss*, trans. by B. Brewster, London: Routledge & Kegan Paul, 97–135

McBrayer, J. (2010a), 'A Limited Defense of Moral Perception', *Philosophical Studies*, 149: 3, 305-320

McBrayer, J. (2010b), 'Moral Perception and the Causal Objection', *Ratio*, 23: 3, 291-307

McDowell, J. (1978), 'On the Reality of the Past', in C. Hookway and P. Pettit (eds.), *Action and Interpretation*, Cambridge: Cambridge University Press, 127-144

McDowell, J. (1979), 'Virtue and Reason', *The Monist* 62: 331-50 (reprinted in McDowell 1998)

McDowell, J. (1996), *Mind and World*, Cambridge, Mass.: Harvard University Press

McDowell, J. (1998), *Mind, Value and Reality*. Cambridge, MA: Harvard University Press.

McDowell, J. (2007a), 'What Myth?', *Inquiry*, 50: 4, 338-351

McDowell, J. (2007b), 'Response to Dreyfus', *Inquiry*, 50: 4, 366-370

McDowell, J. (Unpublished), 'Philosophy of Action', draft made available as part of a workshop on 'John McDowell's Philosophy of Action', at the University of Basel, December 11-13 2009

McGann, M and H. De Jaegher. (2009), 'Self-Other Contingencies: Enacting Social Perception', *Phenomenology and the Cognitive Sciences*, 8: 4, 417-437

McGee, C.D. (1964), 'Fun, Games and Natural Language', *Australasian Journal of Philosophy*, 42: 3, 335-344

McGrath, S. (2004), 'Moral Knowledge by Perception', *Philosophical Perspectives*, 18: 1, 209-228

McNeill, W. (2010), 'On Seeing That Someone is Angry', *European Journal of Philosophy*, Early View, published online 6 July 2010

McPherson, T. (2011), 'Against Quietist Normative Realism', *Philosophical Studies*, 154: 2, 223-240

Mead, G.H. (1934), *Mind, Self and Society from the Standpoint of a Social Behaviourist*, Vol. 1, ed. by C.W. Morris, Chicago: University of Chicago Press

Medina, J. (2010), 'Wittgenstein as a Rebel: Dissidence and Contestation in Discursive Practices', *International Journal of Philosophical Studies*, 18: 1, 1-29

Meltzoff, A.N. (1990), 'Foundations for Developing a Concept of Self: The Role of Imitation in Relating Self to Other and the Value of Social Mirroring, Social Modelling, and Self-Practice in Infancy', in D. Cichetti and M. Beeghly (eds.), *The Self in Transition: Infancy to Childhood*, Chicago: University of Chicago Press, 139-164

Meltzoff, A.N. and A. Gopnik. (1993), 'The Role of Imitation in Understanding Persons and Developing a Theory of Mind', in S. Baron-Cohen, H. Tager-Flusberg, and D.J. Cohen (eds.), *Understanding Other Minds: Perspectives from Autism*, Oxford: Oxford University Press, 335-365

Milgram, S. (1977), *The Individual in a Social World: Essays and Experiments*, London: Addison-Wesley Publishing Company

Millar, A. (2000), 'The Scope of Perceptual Knowledge', *Philosophy*, 75: 73-88

Miller, D.L. (1951), 'Norms, Values, and the Social Sciences', *Southwestern Social Science Quarterly*, 32, 137-149

Miller, S. (2001), *Social Action: A Teleological Account*, Cambridge: Cambridge University Press

Millikan, R. (2005), *Language: A Biological Model*, Oxford: Clarendon Press

Mulligan, K. (1998), 'From Appropriate Emotions to Values', *The Monist*, 81, 161-188

Murakami, K. (2010), 'Liminality in Language Use: Some Thoughts on Interactional Analysis from a Dialogical Perspective', *Integrative Psychological and Behavioural Science*, 44: 1, 30-38

Murray, L and C. Trevarthen. (1985), 'Emotional Regulations of Interactions Between Two-Month-Olds and their Mothers, in T. M. Field and N. A. Fox (eds.), *Social Perception in Infants*, Norwood, NJ: Ablex, 177-197

Murray, L. (1980), *The Sensitivities and Expressive Capacities of Young Infants in Communication with their Mothers*, Edinburgh: University of Edinburgh, PhD Thesis

Nadel, J, I. Carchon, C. Kervella, D. Marcelli, and D. Rtserbat-Plantey. (1999), 'Expectations for Social Contingency in 2-Month-olds', *Developmental Science*, 2, 164-173

Neugebauer, L and D. Aldridge. (1998), 'Communication, Heart Rate and the Musical Dialogue', *British Journal of Music Therapy*, 12: 2, 46-52

Nichols, S. (2004), *Sentimental Rules: On the Natural Foundations of Moral Judgement*, Oxford: Oxford University Press

Noë, A. (2004), *Perception in Action*, Cambridge, Mass.: MIT Press

Opp, K-D. (2001), 'How do Norms Emerge? An Outline of a Theory', *Mind and Society*, 3: 2, 101-128

Ortner, S.B. (1984), 'Theory in Anthropology Since the Sixties', *Comparative Studies in Society and History*, 26: 1, 126-166

Overgaard, S. (2005), 'Rethinking Other Minds: Wittgenstein and Levinas on Expression', *Inquiry*, 48: 3, 249-274

Paglieri, F. (2005), 'Playing by and With the Rules: Norms and Morality in Play Development', *Topoi*, 24, 149-67

Pauen, M. (2012), 'The Second-Person Perspective', *Inquiry*, 55: 1, 33-49

Pettit, P. (1990), 'The Reality of Rule-Following', *Mind*, 99: 393, 1-21

Pinker, S. (1991), 'Rules of Language', *Science*, 253, 530-535

Pinker, S. (1999), *Words and Rules: The Ingredients of Language*, New York: Basic Books

Pinker, S. (2001), 'Words and Rules', *Eye on Psi Chi*, 5: 3, 14-19

Polanyi, M. (1958), *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago: University of Chicago Press

Pollis, N and C. Pollis. (1970), 'Sociological Referents of Social Norms', *The Sociological Quarterly*, 11: 2, 230-242

Putnam, H. (1981), 'Convention: A Theme in Philosophy', *New Literary History*, 13: 1, 1-14

Quinn, M.S. (1975), 'Practice-Defining Rules', *Ethics*, 86: 1, 76-86

Railton, P. (1997), 'On the Hypothetical and the Non-Hypothetical in Reasoning about Belief and Action', in G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Clarendon Press, 53-79 (reprinted in Railton 2003)

Railton, P. (1998), 'Red, Bitter, Good', *European Review of Philosophy*, 3, 67-84

Railton, P. (2000), 'A Priori Rules: Wittgenstein on the Normativity of Logic', in P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori*, Oxford: Oxford University Press, 170-196

Railton, P. (2003), *Facts, Values and Norms: Essays Towards a Morality of Consequence*, Cambridge: Cambridge University Press

Railton, P. (Forthcoming), 'Nietzsche's Normative Theory? The Art and Skill of Living Well', in C. Janaway and S. Robertson, *Nietzsche, Naturalism and Normativity*, Oxford: Oxford University Press

Rakoczy, H, F. Warneken, and M. Tomasello. (2009). 'Young Children's Selective Learning of Rule Games from Reliable and Unreliable Models', *Cognitive Development*, 24, 61-69

Rakoczy, H. (2007), 'Play, Games and the Development of Collective Intentionality', *New Directions for Child and Adolescent Development*, 115, 53-67

Rakoczy, H. (2008), 'Taking Fiction Seriously: Young Children Understand the Normative Structure of Joint Pretend Games', *Developmental Psychology*, 44: 4, 1195-1201

Rawls, A.W. (2004) *Epistemology and Practice: Durkheim's* The Elementary Forms of Religious Life, Cambridge: Cambridge University Press

Rawls, J. (1955), 'Two Concepts of Rules', *Philosophical Review*, 64, 3-32

Raz, J. (1999), 'Explaining Normativity: On Rationality and the Justification of Reason', *Ratio*, 12: 4, 354-379

Reason, J. (1990), *Human Error*, Cambridge: Cambridge University Press

Reber, A.S. (1967), 'Implicit Learning of Artificial Grammars', *Journal of Verbal Learning and Verbal Behaviour*, 6, 855-863

Reck, C, A. Hunt, T. Fuchs, R. Weiss, A. Noon, E. Möhler, G. Downing, E.Z. Tronick, and C. Mundt. (2004), 'Interactive Regulation of Affect in Postpartum Depressed Mothers and Their Infants: An Overview', *Psychopathology* 37, 272-280

Reddy, V. (2008), *How Infants Know Minds*, Cambridge, Mass.: Harvard University Press

Rescorla, M. (2011), 'Convention', *The Stanford Encyclopedia of Philosophy* (Spring 2011 edition), ed. by E. Zalta, http://plato.stanford.edu/archives/spr2011/entries/convention/

Rietveld, E. (2008a), *Unreflective Action: A Philosophical Contribution to Integrative Neuroscience*, Dissertation, University of Amsterdam, Amsterdam: ILLC-Dissertation Series DS-2008-05

Rietveld, E. (2008b), 'Situated Normativity: The Normative Aspect of Embodied Cognition in Unreflective Action', *Mind*, 117, 973-1001.

Rietveld, E. (2008c), 'The Skilful Body as a Concernful System of Possible Actions: Phenomena and Neurodynamics', *Theory & Psychology*, 18: 3, 341-363

Rietveld, E. (2010), 'McDowell and Dreyfus on Unreflective Action', *Inquiry*, 53: 2, 183-207

Rietveld, E. (Forthcoming), 'Context-Switching and Responsiveness to Real-Relevance', in J. Kiverstein and M. Wheeler (eds.), *Heidegger and Cognitive Science*, London: Palgrave

Rifelj, C. (1992), *Reading the Other: Novels and the Problem of Other Minds*, Ann Arbor: University of Michigan Press

Robbins, P and M. Aydede. (2009), 'A Short Primer on Situated Cognition', in P. Robbins and M. Aydede (eds.), *The Cambridge Handbook of Situated Cognition*, Cambridge: Cambridge University Press, 3-10

Robertson, S (ed.). (2009a), *Spheres of Reason: New Essays in the Philosophy of Normativity*, Oxford: Oxford University Press

Robertson, S. (2009b), 'Introduction: Normativity, Reasons, Rationality', in Robertson (ed.) (2009), 1-28

Rochat, P. (2009), *Others in Mind: Social Origins of Self-Consciousness*, New York: Cambridge University Press

Rommetveit, R. (2003), 'On the Role of a "Psychology of the Second Person" in Studies of Meaning, Language and Mind', *Mind, Culture and Activity*, 10: 3, 205-218

Rosch, E. (1978), 'Principles of Categorisation', in E. Rosch and B.B. Lloyd (eds.), *Cognition and Categorisation*, Hillsdale, NJ: Lawrence Erlbaum, 27-48

Rosen, G. (2001), 'Brandom on Normativity, Intentionality and Modality', *Philosophy and Phenomenological Research*, 63: 3, 611-23

Rosen, M. (2001), 'The Role of Rules', *International Journal of Philosophical Studies*, 9: 3, 369-384

Roth, P. (2003), 'Mistakes', *Synthese*, 136, 389-408

Roth, P. (2012), 'Searleword', *History & Theory*, 51, 123-142

Rouse, J. (2007a), 'Practice Theory', in Turner and Risjord (eds.), 639-681

Rouse, J. (2007b), 'Social Practices and Normativity', *Philosophy of the Social Sciences*, 37: 1, 46-56

Ryle, G. (1949), *The Concept of Mind*, London: Hutchinson & Company

Sacks, O. (1985), *The Man who Mistook his Wife for a Hat*, London: Duckworth

Schank, R.C. and R.P. Abelson. (1977), *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*, Hillsdale, NJ: Lawrence Erlbaum

Schear, J. (ed.) (Forthcoming), *Mind, Reason and Being-in-the-World: The McDowell-Dreyfus Debate*, London: Routledge

Schelling, T. (1960), *The Strategy of Conflict*, Cambridge, Mass.: Harvard University Press

Schöner, G and H. Reimann. (2009), 'Understanding Embodied Cognition Through Dynamical Systems Thinking', in Symons and Calvo (eds.), 450-473

Schulte, J. (2007), 'Rules and Reason', *Ratio*, 20, 464–480

Schutz, A. (1967), *Phenomenology of the Social World*, Evanston, IL: Northwestern University Press

Schwyzer, H. (1969), 'Rules and Practices', *Philosophical Review*, 78: 4, 451-467

Searle, J. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press

Searle, J. (1983), *Intentionality: An Essay in the Philosophy of* Mind, Cambridge: Cambridge University Press

Searle, J. (1995), *The Construction of Social Reality*, London: Penguin

Searle, J. (2010), *Making the Social World: The Structure of Human Civilisation*, Oxford: Oxford University Press

Seemann, A. (2008), 'Person Perception', *Philosophical Explorations*, 11: 3, 245-262

Seemann, A. (2010), 'The Other Person in Joint Attention: A Relational Approach', *Journal of Consciousness Studies*, 22, 161-182

Seemann, A. (ed.) (2011), *Joint Attention: New Developments in Psychology, Philosophy of Mind and Social Neuroscience*, Cambridge, Mass.: MIT Press

Sellars, W. (1954), 'Some Reflections on Language Games' *Philosophy of Science* 21, 204-28

Sellars, W. (1997 [1956]), *Empiricism and the Philosophy of Mind*, Cambridge, Mass.: Harvard University Press

Sennett, R. (2008), *The Craftsman*, London: Penguin

Sennett, R. (2012), *Together: The Rituals, Pleasures and Politics of Cooperation*, London: Allen Lane

Sibley, F. (1955), 'Seeking, Scrutinising and Seeing', *Mind*, 64: 256, 455-478

Sibley, F. (1959), 'Aesthetic Concepts', *The Philosophical Review*, 68: 4, 421-450

Sibley, F. (1965), 'Aesthetic and Non-Aesthetic', *Philosophical Review*, 74, 135-159

Siegel, H (ed.). (2009), *Oxford Handbook of Philosophy of Education*, Oxford: Oxford University Press

Silva, J. (1981), 'Normative Compliance and Rule-Violating Behaviour in Sport', *International Journal of Sport Psychology*, 12, 10-18

Skorupski, J. (2007), 'What is Normativity?', *Disputatio*, 2: 23, 247-269

Smith, B. (2006), 'Particularism, Perception and Judgement', *Acta Analytica*, 21: 2, 12-29

Smith, E, C. Langston and R. Nisbett. (1992), 'The Case for Rules in Reasoning', *Cognitive Science*, 16, 1-14

Smith, J. (2010), 'Seeing Other People', *Philosophy and Phenomenological Research*, 81: 3, 731-748

Smith, L. (2006), 'Introduction, in Smith and Vonèche (eds.), 1-35

Smith, L. and J. Vonèche (eds.). (2006) *Norms in Human Development*, Cambridge: Cambridge University Press

Southwood, N. and L. Eriksson. (2011), 'Norms and Conventions', *Philosophical Explorations*, 14: 2, 195-217.

Starkey, C. (2006), 'On the Category of Moral Perception', *Social Theory and Practice*, 32: 1, 75-96

Stawarska, B. (2009), *Between You and I: Dialogical Phenomenology*, Athens: Ohio University Press

Steiner, P and J. Stewart. (2009), 'From Autonomy to Heteronomy (and Back): The Enaction of Social Life', *Phenomenology and the Cognitive Sciences*, 8: 4, 527-550

Stern, D. (2004), *The Present Moment in Psychotherapy and Everyday Life*, New York: W.W. Norton

Stout, R. (2010), 'Seeing the Anger in Someone's Face', *Aristotelian Society Supplementary Volume*, 84: 1, 29-43

Stueber, K. (2004), 'How to Think about Rules and Rule Following', *Philosophy of the Social Sciences*, 35: 3, 307-323

Stueber, K.R. (2006), *Rediscovering Empathy: Agency, Folk Psychology, and the Human Sciences*, Cambridge, Mass.: MIT Press

Symons, J and P. Calvo (eds.), *The Routledge Companion to Philosophy of Psychology*, Abingdon: Routledge

Tanney, J. (2000), 'Playing the Rule-Following Game', *Philosophy*, 75: 292, 203-224.

Tappolet, C. (2000), *Émotions et Valeurs*, Paris: Presses Universitaires de France

Thelen, E, G. Schöner, C, Scheier, L.B. Smith. (2001), 'The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching', *Behavioral and Brain Sciences*, 24, 1-86

Thomson, J.J. (2008) *Normativity*, Chicago: Open Court

Toulmin, S. (2003), *Return to Reason*, Cambridge, Mass.: Harvard University Press

Travis, C. (2004), 'The Silence of the Senses', *Mind*, 113: 449, 57-94

Tronick, E.Z. and J.F. Cohn (1989), 'Infant-Mother Face-to-Face Interactions: Age and Gender Differences in Coordination and the Occurrence of Miscoordination', *Child Development*, 60, 85-92

Tronick, E.Z. and M.K. Weinberg. (1997), 'Depressed Mothers and Infants: Failure to Form Dyadic States of Consciousness', in L. Murray and P.J. Cooper (eds.), *Postpartum Depression and Child Development*, New York, Guilford, 54-84

Tubert, A. (2010), 'Constitutive Arguments', *Philosophy Compass*, 5: 8, 656-666

Turiel, E. (1983), *The Development of Social Knowledge: Morality & Convention*, Cambridge: Cambridge University Press

Turner, J.H. (2002), *Face to Face: Toward a Sociological Theory of Interpersonal Behaviour*, Stanford: Stanford University Press

Turner, S.P. (1994), *The Social Theory of Practices: Tradition, Tacit Knowledge and Presuppositions*, Cambridge: Polity Press

Turner, S.P. (2007a), 'Explaining Normativity', *Philosophy of the Social Sciences*, 37: 1, 57-73

Turner, S.P. (2007b), 'Practice Then and Now', *Human Affairs*, 17: 2, 111-125

Turner, S.P. (2010). *Explaining the Normative*, Cambridge: Polity Press

Turner, S.P. (Forthcoming), 'Davidson's Normativity', in J. Malpas (ed.), *Dialogues with Davidson: On the Contemporary Significance of His Thought*, Cambridge, Mass.: MIT Press

Turner, S.P. and M. Risjord (eds.). (2007), *Handbook of the Philosophy of Science. Volume 15: Philosophy of Anthropology and Sociology*, Dordrecht: Elsevier

Turner, V. W. (1969), *The Ritual Process: Structure and Anti-Structure*, New York: Aldine de Gruyter

Tye, M. (2005), 'On the Non-Conceptual Content of Experience', in M.E. Reicher and J.C. Marek (eds.), *Experience and Analysis*, Vienna: Proceedings of the 27th International Wittgenstein Symposium, 221-239

Varela, F and N. Depraz. (2005), 'At the Source of Time: Valence and the Constitutional Dynamics of Affect', *Journal of Consciousness Studies*, 12: 8-10, 61-81

Vaz, E. (1977), 'Institutionalised Rule Violation in Professional Hockey: Perspectives and Control Systems', *Journal of the Canadian Association for Health, Physical Education and Recreation*, 43: 3, 6-14

Velleman, D. (1996), 'On the Possibility of Practical Reason', *Ethics* 106: 694-726

Velleman, D. (2009), *How We Get Along*, Cambridge: Cambridge University Press

Von Wright, G.H. (1963), *Norm and Action: A Logical Inquiry*, London: Routledge & Kegan Paul

Watkins, M and K.D. Jolley. (2002), 'Pollyanna Realism: Moral Perception and Moral Properties', *Australasian Journal of Philosophy*, 80: 1, 75-85

Wedgewood, R. (2001), 'Sensing Values?', *Philosophy and Phenomenological Research*, 63: 1, 215-223

Weinberg, M.K. and E.Z. Tronick. (1998), 'The Impact of Maternal Psychiatric Illness on Infant Development', *Journal of Clinical Psychology*, 59, 53-61

Weinberger, O. (1985), 'Freedom, Range for Action, and the Ontology of Norms', *Synthese*, 65, 307-324

Weinrich, B. (1988), *Yiddish Folktales*, trans. by L. Wolf, New York: IVO Institute for Jewish Research

Wheeler, M. (2005), *Reconstructing the Cognitive World: The Next Step*, Cambridge, Mass.: MIT Press

Whiting, D. (2012), 'Are Emotions Perceptual Experiences of Value?', *Ratio*, 25: 1, 93-107

Wide, S. (2009), 'On the Art of Being Wrong: An Essay on the Dialectic of Errors', *Journal of the Philosophy of Education*, 43: 4, 573-588

Williams, M. (2009), *Blind Obedience: The Structure and Content of Wittgenstein's Later Philosophy*, London: Routledge

Williamson, T. (2000), *Knowledge and Its Limits* (Oxford: Oxford University Press)

Winnicott, D.W. (2005 [1971]), *Playing and Reality*, London: Routledge

Wittgenstein, L. (1953), *Philosophical Investigations*, trans. by G.E.M. Anscombe, London: Blackwell.

Wittgenstein, L. (1958), *The Blue and Brown Books*, Oxford: Oxford University Press

Wittgenstein, L. (1967), *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief*, Berkeley: University of California Press

Wittgenstein, L. (1979), *On Certainty*, trans. by G.E.M. Anscombe and D. Paul, Oxford: Blackwell

Wittgenstein, L. (1980), *Culture and Value*, trans. by P. Winch, Oxford: Basil Blackwell

Wittgenstein, L. (1983), *Remarks on the Foundations of Mathematic*s, G.H. von Wright and R. Rhees (eds.), trans. by G.E.M. Anscombe, Cambridge, Mass.: MIT Press

Wright, C. (1980), *Wittgenstein on the Foundations of Mathematics*, London: Duckworth

Wright, C. (2007), 'Rule-Following Without Reasons: Wittgenstein's Quietism and the Constitutive Question', *Ratio (New Series)*, 20, 481–502

Wyman, E, H. Rakoczy, and M. Tomasello. (2009). 'Normativity and Context in Young Children's Pretend Play, *Cognitive Development*, 24, 146-155

Zahavi, D. (2010), 'Empathy, Embodiment and Interpersonal Understanding: From Lipps to Schutz', *Inquiry*, 53: 3, 285-306

Zalabardo, J. (2009), 'One Strand in the Rule-Following Considerations', *Synthese*, 171, 509-51