Self-in-a-Vat: On John Searle's
Ontology of Reasons for Acting

LAURENCE KAUFMANN
University of Lausanne, Switzerland

John Searle has recently developed a theory of reasons for acting that intends to rescue the freedom of the will, endangered by causal determinism, whether physical or psychological. To achieve this purpose, Searle postulates a series of "gaps" that are supposed to endow the self with free will. Reviewing key steps in Searle's argument, this article shows that such an undertaking cannot be successfully completed because of its solipsist premises. The author argues that reasons for acting do not have a subjective, I-ontology but a first-person plural, We-ontology that better accounts for agency and responsibility.

Keywords: free will; agency; reasons for acting; ontology

John Searle always raises fundamental issues for philosophers and sociologists.[1] He has raised questions about language (how can meaning be passed on by mere sound waves; Searle 1969), about the mind (how the mind can be connected to the physical structure of the brain; Searle 1992), and about society (how institutions can be part of an objective world; Searle 1995). In *Rationality in Action*, Searle (2001b) now wonders how individuals can escape from the causal determinism of natural forces to freely act upon reasons. He attempts to clarify and refocus the debate by using a conceptual apparatus cutting across the interdisciplinary boundaries separating the philosophies of mind, language, and society (Clément & Kaufmann, 1996). According to Searle, the common building blocks of the human sciences are intentionality, that is, the property for a mental state to be *about* something, and speech acts. These are the only phenomena whose ontol-

ogy fits the latest discoveries of the exact sciences, that is, the theory of
biological evolution of species and the theory of physical particles of
matter. However, as Searle shows it, this common naturalist ontology
runs into numerous difficulties. By relying on the impersonal, third
person viewpoint peculiar to the science, it makes so-called first per-
son phenomena, such as consciousness, free will, rational decision
making, and reasons for acting, extremely difficult to account for.
How indeed can an externalist, objectivist approach describe what
seems to be submitted to a subjective ontology, that is, a mind-
dependent ontology? The starting point of Searle's inquiry is this
apparent incompatibility between the mental superstructure of the
mind, which would be the homeland of free will and subjective deci-
sion making on one hand, and the material infrastructure of the brain,
which is causally determined, on the other. The main question con-
cerning free will is indeed "How can it be the case that we perform
free actions, given that every event has a cause and causal determina-
tion make free actions impossible?" (Searle 2001b, 4).

Actually, for Searle, deterministic physical mechanisms do not
endanger the freedom of the will because reasons for acting are not
unconscious causes, neurobiologically implemented in the brain.
Against what he calls the "Classical Model of Rationality," he claims
that ends, motivations, beliefs, and desires do not function as both
causes and reasons for acting.[2] If it were the case, the agent could not
do otherwise than acting upon his prior set of beliefs and desires, and
the sense of freedom in voluntary action would be a mere illusion. In
Searle's framework, reasons for acting are far from being a fixed set of
preferences that would function as sufficient conditions for an action
to occur. On the contrary, reasons are nothing but conflicting and
incompatible ends that require evaluation and deliberation on the
agent's part (Searle 2001b, 30-31). Thanks to the double "gap" that
separates reasons for acting and the final action—that is, a gap
between reasons and decision making and a gap between the decision
and its carrying out—ends, beliefs, and desires are causally ineffi-
cient. This claim has two important implications for Searle's
approach. First, in virtue of those gaps, the agent's power to knowl-
edgeably choose his actions is safeguarded: rational agency and free

2. According to Searle, the original sin of such a model consists in reducing rational
decision making to the selection of means necessary for achieving one's desired ends.
In accordance with Kant's famous claim, that is, "he who wills the ends wills the
means," the so-called Classical Model of Rationality wrongly sees deliberation as being
always about means, never about ends (Searle 2001b, 4).

will depend on the existence of an irreducible, conscious *self*, which has the capacity to make a reason effective by acting on it. Second, practical reason is different from theoretical reason, so emphasized in the so-called "Classical Model of Rationality," which would provide the agent with the rules of logical inferences enabling him to select the means necessary for achieving his desired ends. For Searle, "rationality in action" consists indeed in finding some way to adjudicate between various inconsistent aims and to give up some desires in order to satisfy others.

In this article, I will leave aside Searle's theory of rationality to discuss only two aspects of his philosophy, both concerned with ontological issues. In the first part of my argument, I will discuss the issue of causal determinism and Searle's solution to the opposition between the "indeterminate" level of the mind, submitted to the free will of agents, and the "determinate" level of the brute, blind forces of the unconscious bits of matters that make up the brain. I will argue that this alleged solution is indeed very problematic because it mainly relies on what could be called "the multifaceted causation" of "biological naturalism" (part I). In the second part of my argument, I will focus on the ontology that underlies Searle's conception of reasons for action, with special emphasis on his distinction between two "species of reasons": the traditional motivational set of desires and needs, shared by other evolved primates, which refers to things agents *want* to do, and the second species of reasons, specifically human, which refers to things agents *have to* do, whether they want to or not (part II). I will argue that the concept of desire-independent reasons, although insightful and promising, does not come up to the expectations it might raise, mainly because of its solipsist, individualistic premises. After an ontological reconstruction of those individualistic grounds, I will propose an alternative approach to reasons for action, more compatible, in my sense, with their social and institutional dimension (III).

I. BIOLOGICAL NATURALISM

According to Searle, agency logically requires a conscious entity, namely, the self, which has the capacity to initiate and carry out actions. But it has still to be shown, Searle says, that the self, free will, and the gap between reasons and action are genuine ontological phenomena and not illusions: the supposed agentive behavior might

indeed be entirely preprogrammed with the physical deterministic mechanisms ruling the brain. In this case, conscious decision, free will, and the self would have no causal power as such and would therefore be *epiphenomenal*. Although agents feel they have power and freedom when they act, their actions would be determined by unconscious causes (Searle 2001b, 280-81). To avoid this deterministic conclusion, naturalist philosophy has to find a way to account for free will without infringing the requirements of the materialist, realist ontology of the physical sciences. That is precisely the challenge Searle tries to take up, hence his commitment to monist ontology (Ia), to the efficient causation that is supposed to underlie it (Ib) and to methodological solipsism (Ic).

Ia. A Monist Ontology?

Against a tradition that wrongly divides the world into two forms of reality, the mental and the physical, Searle maintains that we live in a single world. Nevertheless, in this unique world, there are two kinds of phenomena, the *physical*, which exist independently of us, and the *mental*, which exist only relative to our intentionality (Searle 1998, 2000). To prevent a dualistic ontology, split up into physical and mental phenomena, brain mechanisms and the contents of consciousness have to be related by a bottom-up causality (Searle 2000, 566). This causal reductionism is indeed essential for a unitary naturalist ontology because it permits at any time the reduction of high-level mental states to the low-level neurophysiological goings-on in the brain. Such a reductionism ensures the material identity between the empirical referents of neurological and mental states and hence their strict correspondence in the world independently of any observer's perspective (Colin 1997, 230-32). In the same way that the extension of the concept of water can be empirically replaced by its molecular formula H_2O, the extension of mental states, whether conscious or not, are replaceable by their underlying neurological substrate at each stage of their development. Against any version of the dualism of properties, which holds that physical implies nonmental, Searle's biological naturalism claims that mental and physical properties are not mutually exclusive: some physical properties are at some level of description mental whereas all mental properties at another level of description are physical (Corcoran 2001b, 311).

In Searle's framework, however, this causal reductionism is in no way a kind of ontological reductionism (Searle 1991, 182; Searle 1992).

While insistently defending a single natural ontology, Searle endows consciousness with a specific "subjective ontology." This is quite surprising coming, as it does, from someone seeking to refute any dualistic ontology as a throwback to a traditional, obsolete metaphysics (Searle 1998, 42). From the viewpoint of Searle's subjective ontology, the private, hidden, internal, and subjective mind is indeed opposed to the public, observable, external, and objective brain. As Olafson (1994) put it, this theory entails a kind of "brain dualism" that isolates a subjective brain from an objective one.[3] Although mental states are caused by brain processes and are nothing more than biological entities, they are ontologically irreducible and accessible only to the inner consciousness. Searle's oscillation between, on one hand, a monist ontology, which is merely provisional, assumed rather than defended, and, on the other hand, the distinctive existence of the mental, thus makes micro-macro causation a crucial point. Micro-macro causation indeed is the only remaining ontological link able to prevent mental phenomena from an autonomization that would definitely be synonymous with a parallel, immaterial world. In other words, causation has a central role in Searle's cosmology, in that it is designed to come to the rescue of a monism jeopardized by the potential incompatibility between the mental and physical accounts of the reality. To the extent that micro-macro causation is the last connection that Searle has at his disposal to fill in the dangerous gaps menacing his alleged unified ontology, we need to sketch out the main ways he uses it.

Ib. An Erratic Efficient Causation

Searle claims that the attempt to distinguish between different kinds of causation is wrong and superfluous: there is only one kind of causation, and that is efficient causation (Searle 1983, 135; Searle 2001b, 29). As everyone knows, the paradigm of efficient causation is that of billiard balls: a billiard ball provokes the movement and hence the trajectory of another one by bumping into it. The causal link connecting the antecedent and consequent events is physical and implies

3. Despite his rhetorical precautions and reiterated rejection of all kinds of dualism, Searle takes up the dualism between subjective and objective, first-person and third-person phenomena. His approach implies de facto a bifurcation of biological properties of the brain, some properties being accessible to any surgeon whereas the remainder is accessible only to consciousness. As Corcoran (2001, 314) put it, Searle's biological-property dualism has thus "all the advantages of theft over honest toil."

the existence of two a priori separated entities, in this case two billiard balls. Strangely, for Searle (2001b 41), intentional causation, which refers to the cases in which a state of affairs causes a mental state (i.e., the frog on the water lily creates my visual experience) or a mental state causes a state of affairs (i.e., the intention to eat frogs causes the fact that I eat frogs), is a subcategory of efficient causation. Even if the efficient causation in question does often switch from the "push-pull" paradigm to the "making something else happen" one, the basic pattern remains the same (Searle 1998, 59).

In my own discussion below, I would like to show that the generic status of efficient causation, duly in charge of the unification of an ontology that risks collapsing at any time, is obviously refuted by the three mysterious shapes that this so-called efficient causation takes in Searle's work. In my sense, the three kinds of causation I am going to briefly pull apart are too mismatched to play the role of ontological glue they were supposed to.

Ib.1. Conscious mental states, which are supposedly "*caused* by the operations of the brain and *realized* in the structure of the brain" (Searle 1983, 265), resort to what could be called Searle's first kind of causation. The emergence of conscious macrophenomena, and then of their higher order properties, is the direct effect of the activation of neuronal microphenomena, just like the liquidity of water emerges from the formation of H_2O molecules. To the extent that consciousness and intentional states are nothing but high-level states of the cerebral system, their reality is not distinct from the low-level reality of neurons they are made of. For Searle (1991, 181-85), the causal effect of neuronal micro-states, often called "supervenience," consists in generating mental macro-properties that are, in turn, causal—provided that they escape in one way or another from the physical deterministic mechanisms ruling the brain. The problem, according to me, is that the brain-mind causation in question here, if any, is obviously not the external relationship between two distinct and independent entities peculiar to efficient causation. It is rather an internal connection between two codependent phenomena, namely, mind and brain, which turn out to be different levels of the same whole structure. This codependence is then ontological and not causal: as Esfeld (1998, 367) pointed out, generic ontological dependence among the parts of a holistic system means that there can be no occurrence of a certain type unless there is some occurrence of another specific type. But when the effect (the mind) is so dependent on the cause (the brain), it is not

about efficient causation anymore; both are parts of a unique whole system whose different levels are determined from "below," that is, from the neurological circuits of the brain. Moreover, if consciousness is the top level of a whole system, deterministic on every level, then conscious decisions are devoid of any "top-down" causal power. The analogy between mental properties and other supervenient properties such as liquidity or solidity is therefore misleading: whereas liquidity or solidity can be defined by their specific causal powers, conscious phenomena are defined by their subjective, experiential quality (Nida-Rümelin, 2002). The causal impotence of the first-person features of conscious mental states thus puts free will in jeopardy.

Ib.2. Partly aware of this tremendous difficulty, Searle introduces, in *Rationality in Action*, what could be called his second kind of causation in order to save psychological indeterminism required by free will from material determinism. Like Penrose (1994), he points to the nondeterministic physical processes that supposedly prevent the neurobiological underpinnings of psychology from being causally sufficient. Those processes, ruled by the indeterminacy of quanta mechanics, leave a gap between what is happening now in the conscious volitional component of the whole-brain system and what is going to happen next (Searle 2001b, 287-88). The state of the brain-mind system at a given time is not causally sufficient to determine the next state of the system. In other words, for Searle, the lack of causally sufficient conditions at the psychological level goes all the way down to the lowest level of causal determination. It does not stop at the level of neurons but continues its descent until the quantum mechanical level. In virtue of the "underdeterminacy" of psychological states by their physical substrate, consciousness in general and decision making in particular could causally affect each neuron of the system they are part of. And free will would have an important place in the system.

Unfortunately, this kind of quantum causation, just like the bottom-up link mentioned above, turns out to be an internal connection that does not deserve as such the label of efficient causation. Moreover, in my view, such a causation is surprisingly selective. How indeed can one explain that all high-level properties of the mental (e.g., subjectivity, qualitativeness, and unity) are distinguishable from low-level properties, except for the property of indeterminacy, which would be miraculously preserved cross level? A causal link is

supposed to be systematic; it cannot transfer only one unchanged neuronal property, namely, indeterminacy, to the upper level of the mind while drastically transforming the remaining brain properties into subjective features. Moreover, even if such a selective causation were conceivable and mental phenomena were effectively indeterminate, it would not necessarily endow conscious decision making with the final determination of the action. On one hand, this argument commits a genetic fallacy; it is not because the elementary building blocks at the origin of mental phenomena are indeterminate in nature that these mental phenomena are themselves indeterminate. After all, billiard balls are also grounded on indeterministic quanta mechanisms, and this fact does not prevent their shapes, properties, and trajectories from being perfectly determinate and predictable. On the other hand, the indeterminacy of mental phenomena, if any, does not necessarily need to be offset by the decisive intervention of the self. Logically, the final determination of the action could be carried out by other mechanisms as well, for instance, the nonintentional workings of abilities, dispositions, and skills or the behavioral triggers provided by the external world (Ryle 1949/1984).

Ib.3. Apart from the (too) intimate relationships between cause and effect that characterize the brain-mind causation, Searle implicitly assumes another kind of so-called efficient causation. What could be called his third kind of causation leaves the cross-level of micro-macro relations and takes place in the upper level of mental states. It indeed relates reasons for acting (desires, motivations), intentions or decisions, and the effective action "*modulo the gap*" that separates each of them (Searle 2001b). Thanks to this gap, the causal link between the reasons for acting and the effect that the action itself constitutes is necessary but not sufficient. The "self," which has by definition a full range of alternatives available to it, adds the subjective, free, and decisive surplus sufficient to bring about the action. Even if Searle (2001b, 17) understandably rejects this way of putting it by claiming that "nothing fills the gap, you just make up your mind to do something," the final causal power, denied to any other stages of decision making, is de facto bestowed upon the mysterious figure of the self. The self is duly in charge of the final selection of the reason for acting. Although the self is a mere "formal requirement on rational action" and a mere logical property of the "single unified conscious field" (Searle 2001b, 93-95), it thus enjoys a causal power.

This bizarre conclusion does not seem unique in the work of Searle, who openly claims that intentional states and constitutive rules "function *causally* in virtue of their *logical* structure" (Searle, 2001a, 278). Such confusion between the logical level of what must theoretically be the case and the ontological level of what causally occurs is more than embarrassing. Although the logical inquiry is an oblique way to reach the ontology, it belongs to a formal mode of description of the reality, not to the reality itself (Récanati 1984). For instance, human behaviors can be described in terms of rules without being effectively governed by these rules. The assertion that the analytical structure aiming at describing the world is causal as such amounts to the claim that "the roads must be red because they correspond to red lines on a map" (Ziff 1960, 38). By giving a causal power to logical structures, Searle mistakes the model of the reality for the reality of the model. Now, as will be seen, his model of the self does not need this supplementary difficulty: it is already litigious enough.

Ic. A Solipsist Stance

At the end of this quick review, it appears that the disparate versions of the alleged efficient causation Searle claims to use are anything but efficient.[4] Whether it be the "supervening" bottom-up causation and the indeterministic quanta causation that characterize the micro-macro relations or the causal power of the logical entity of the self that characterizes the intentional level, none of them really matches either the "push-pull" paradigm or the "making things happen" model. Those eclectic subspecies of efficient causation do not indeed seem to be able to perform the function they were entrusted with, namely, bridging the potential gaps that separate the different levels of complexity of living matter. In other words, those allegedly efficient causal links are too mismatched and ad hoc to cement together the different layers of Searle's ontology: the absence of genu-

---

4. In fact, one could still find a fourth mysterious subspecies of causation in Searle's work. This causation concerns what he calls the "pre-intentional background" of expectations, procedures, and capacities, due to both human biological equipment (e.g., vision, audition, speech) and local cultural customs (e.g., to pay with money, to take the bus, to eat frogs; Searle 1983, 1995). The background causal role consists merely of a weak and not decisive causation because "it enables without determining intentional contents" (Searle 1983, 158). But we will get back later to the issue of the background and its rather puzzling kind of efficient causation.

ine causal workings makes the monism of his "biological naturalism" appear more a matter of wishful thinking than a matter of fact.

Once Searle's pseudo-solution to the mind-body problem is shown to be empty, we have to see whether or not he does better in the second part of his undertaking. This latter is concerned, so to speak, with the upper ontological layer of his model, namely, the ontology of reasons. In my sense, Searle's ontology of reasons is also marked by a strong dualism but a dualism that is no longer bottom-up, such as the brain-mind connections. This dualism is inner-outer because it radically separates internal minds from the external world. Actually, the psychological world in which Searle's sovereign self reigns is almost impervious to the outer world, including others' minds. To me, this odd imperviousness is the consequence of Searle's "methodological solipsism," that is, as Putnam (1975, 220) put it, "the assumption that no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed." According to this radical individualist perspective, the brain has a narrow mental life and can function according to an internal causal circuit, as the thought experiment of the "brain in a vat" shows. For Searle (1983, 187), an isolated brain in a vat, fed only on information transmitted by some crazy scientists, could develop a perfectly high-functioning mental world. In this virtual brain, just like in normal brains, intentional states can be defined and explained independently of any reference to their physical or sociolinguistic environment. So, "I could have all the intentionality I do have even if I am radically mistaken, even if the apparent presence and cooperation of other people is an illusion, even if I am suffering a total hallucination, even if I am a brain in a vat" (Searle 1990, 407). In other words, no contribution of the outside world is needed for the determination of intentional states and, a fortiori, of behaviors. The inner world, fundamentally self-sufficient, relies only on the internal states of the brain. For Searle, we could therefore all be "brains in a vat." After the inputs related to the person's insertion in the external world have been either deduced from her psychological states or translated into representational contents, the only things that really matter happen *in her head*. Admittedly, the solipsist dogma does not mean the ontological disappearance of the external world. It just aims at simplifying the methodology by doing as if agents were structurally alone in the world. But like any method, methodological solipsism has wide, and not necessarily valuable, heuristic consequences. That is what Searle's ontology of reasons shows, as I will try to demonstrate by reconstructing,

step by step, his reasoning. To do so, I will start with comments on the general status of reasons for acting (IIa) before analyzing more precisely the concept, potentially crucial for the social sciences, of desire-independent reasons (IIb).

II. THE ONTOLOGY OF REASONS

IIa. A World of One's Own?

If we are to believe Searle, the reason for any given behavior is the reason that is subjectively intended by the agent. "Why is it a reason *for me*? Because I have freely created it as a reason for me" (Searle 2001b, 189). Paradoxically, this reason can be immoral (e.g., to exploit others for one's own profit) or irrational (e.g., to persist in smoking while knowing that this can provoke cancer some years later), but its status of reason depends only on the self's "recognition." Actually, only the agent's "recognitional rationality," that is, the representation of reasons by some internal intentional states, is able to elect some motivators as providing a valid reason for acting.

Although sensible at first sight—after all, reasons have to be recognized by the agent to become effective—this claim turns out to be very problematic. In fact, what deserves or not the label of reason is not a mere matter of philosophical vocabulary. Endowing, like Searle's model, any irrational, immoral, or absurd motives with the status of reasons for acting dismisses a constitutive property of reasons, namely, the ability to justify what happened. As Winch (1958, 81) put it, a husband who says he killed his wife out of jealousy *explains* his behavior but he does not *justify* it because such a motive does not meet the "standards of reasonable behavior current in society." Admittedly, if one replaces the misleading concept of reasons in Searle's theory with the notion of motives, one could argue, with Searle, that public and acceptable reasons have nothing to do with effective motives. Good, justifiable reasons would have no role to play in the explanation of action because their use would be limited to the superficial, polite, and conventional way to deal with one another in the public domain. For instance, a well-mannered individual will invoke a pretext, say a cold, to decline an invitation to dinner although the real reason of his refusal is in fact the bad company of his hosts (von Wright 1983). In this case, the motivating, effective reason is different from the good, justifying reason.

This being so, the distinction between good and effective reasons must not be overestimated. If hypertrophied, it risks relegating all good reasons for acting to the shallow range of excuses or pretexts. That is precisely what Searle tends to do: to him, good reasons indicate only why an action *should* have been done and not the effective motives stating why it was *in fact* done (Searle 2001b, 110). Yet it is not certain that the motivational and normative components of reasons for acting are as strictly separated from one another as this dualistic conception suggests. Actually, such a separation tends to draw an impervious boundary between what is "reasonable" to do for a member of a given community and what is "rational" to do for a subject relative to his practical selection of ends and means. This hermetic boundary is mainly due to Searle's definition of rationality as a quality of the reasoning process, namely, the inner coordination of one's own beliefs, desires, hopes, fears, perceptions, and other intentional states (Searle 2001b, 23). Now, this subjective, mentalist conception of rationality comes up against an essential requirement for rational behavior: the evaluation of what would be *objectively reasonable* for anyone to perform. A person who has ordered a beer in a bar can technically refuse to pay for it, but this refusal, unlike Searle's account of it, manifests less the inalienable exercise of her free will than the irrationality of a decision literally inconsequential. This refusal is indeed completely irrational from the angle of its logical, unpleasant consequences, such as the insults of the waiter, the public disapprobation of other customers, or the troubles with the police.

To be genuinely rational, a practical agent cannot thus content himself with coordinating his ends as a function of his desires and beliefs in the inner space of his head. He has to anticipate and evaluate what could and should be objectively reasonable, justifiable, and appropriate to do given the circumstances, the foreseeable reactions of others, and his own capacities. As MacIntyre (1986) points out, a person who greedily devours the unique exemplar of a fruit designed to reduce the hunger in the world and gives her own appetite as the reason for her action would appear irrational, asocial, and immoral. In other words, the glutton's idiosyncratic motive for eating the precious fruit does not deserve the public status of a genuine reason for acting.

And yet that is precisely the incongruous conclusion that Searle's attempt to ground reasons for acting in the subjective attitudes of solitary individuals leads to. By giving absolute priority to the subjectively intended reasons, his model dismantles the impersonal reasons endowing actions with an objective meaning. As a result, one can call

someone all sorts of names or sexually harass a woman while claiming that the reasons of one's action have nothing to do with insulting or assaulting. Similarly, in Searle's model, the fact that a woman is in trouble is not an objective reason to help her because *it is up to the agent* to subjectively consider this fact as compelling or not. By making the determination of action totally independent of its likely interpretation by the community of reasonable beings, Searle's solipsist view circumvents the objective, impersonal, and public reasons that are embodied in the actions in question. Such a conception endangers the commonality of reasons and opens up an "impracticable" social world, one that is no longer stable, shared, and shareable in common with others.

In my sense, to avoid this unlikely world, one has to give up conceiving reasons for acting as mental objects, determined in the mind before any "acting out." In contrast to Searle's alleged gap between reasons for acting and the action itself, fulfilled with the idiosyncratic and empirical work of conciliation conducted by the self, I think that reasons are mostly internally linked to actions. Actually, as Mills (1940) put it, every meaningful action is learnt and associated with a public "vocabulary of motives" from which private purposes must derive to count as valid reasons for acting. Reasons-giving is a capacity of standing under a general principle that one's actions are justified (i.e., rationally valid) only if they could meet with the agreement of others (Baynes 2001). This capacity being "as little in my mind as the movements are in my body" (Ricoeur 1977, 93), reasons are not, contrary to Searle's claim, "factitive entities" whose propositional structure would entail a "that-clause" allowing the self to adopt a relatively distal attitude with regard to them. That is only under the auspices of theoretical rationality that reasons for acting might appear as mental sentences, intrinsically propositional, summoned before the tribunal of the consciousness. As for practical reasons, they are not something that individuals have in their possession. Practical reasons are constitutive of the common, objective, public intelligibility of actions and thereby enjoy a pragmatic status. Whatever one's personal motives may be, to raise one's arm in an official assembly aims to ask for permission to speak, to call someone a rotten bastard means an insult, and to order a beer is a commitment to pay for it. In those cases, one cannot claim "I didn't mean it" when asked for the reasons for one's action. Unless the agent is a cultural dope or an ontological solipsist confined in his own world, he is capable of endorsing the range of reasons, entitlements, and commitments recognized as appropriate by his community. But to make this case further, we need

to dwell on an essential element of Searle's argument, namely, the status of desire-independent reasons.

IIb. Desire-Independent Reasons

According to Searle (2001b), what he calls the Classical Model of Rationality wrongly makes desires the primitive and unique drive to act. In so doing, such a model forgets that desires are not the generic matrixes of any reason for acting. Actually, Searle says, human beings are characterized by the ability to create, recognize, and act on "*desire-independent reasons*" for action, such as obligations, commitments, responsibilities, or requirements. For instance, if I promise to come to your party, I have the obligation, and so the desire-independent reason, to come to your party. For Searle, commitments and obligations have a binding force because rational beings tend to recognize commitments and obligations as decisive reasons for acting. This being so, this "recognitional rationality" does not endanger free will. Agents are not necessitated but motivated by their commitments and obligations, mainly thanks to the irreducible "gap" between the recognition of those commitments and obligations as valid reasons for acting and the final making of the action.

In my sense, Searle's "desire-independent reason" is a handy, insightful concept, which promises to bridge the eternal gap that separates, in philosophy as well in the social sciences, intentional states and institutional facts, the mental universe and the social world. However, as I will try to show, the individualist bias of Searle's approach tremendously weakens the heuristic potential of such a concept. To highlight the logical mistakes at the origin of this failure, I will split up, for clarity's sake, Searle's reasoning on desire-independent reasons into four different episodes. In the first episode, the agent binds freely his will; by ordering a beer, he creates the obligation to pay for it (IIb.1). In the second episode, the agent's obligation becomes an *is* and hence an "epistemically objective fact in the world"; the obligation he creates to pay for his beer is so to speak *out there* (IIb.2). In the third episode, the subjective recognition, by the agent, of that *is* as being a valid reason for action turns it into an *ought*; he knows he has to pay for this beer, whether he wants to or not (IIb.3). In the fourth episode, however, the agent still has the choice to decide to effectively act upon this *ought*; in virtue of his free will, he can refuse to pay for his beer. The supposedly genuine desire-independent episode, that is the second one, is thus surrounded by three episodes driven by free will

and monitored by the self. We have already dealt with the fourth episode. Even though the agent can technically refuse to pay for his beer, such a decision is less the expression of a genuine freedom than the mark of an error of judgment by a "judgmental dope." So let us analyze the underlying assumptions of the three remaining episodes.

IIb.1. First Episode: The Self-Creation of Norms

Oddly enough, in Searle's model, obligations and commitments are similar to desires and intentions: they supposedly respond to the same direction of fit, namely, an "upward," world-to-mind direction of fit. Just as a given intention (i.e., the intention to eat a cake) can be satisfied only by the effective realization of the intended state of affairs (i.e., actually eating a cake), a given commitment (i.e., the commitment to go to Chicago) would be satisfied only if the world really is the way the agent says it is or will be (i.e., I actually go to Chicago). This rather counterintuitive direction of fit of obligations and commitments is due to the fact that, in Searle's view, obligations and commitments are the logical consequences of the willing self-involvement on the part of the agent, who decides to commit himself and to bind thereby his will in the future. The decision of committing and obligating oneself is the prior ground of subsequent, and sometimes undesired, consequences. For Searle, although the self is supposed to face those undesired consequences, it can always choose to act upon one's inclinations instead of meeting one's commitments.

From my point of view, the problem is that this conception does not take seriously the meaning of "*desire-independent* reasons." By definition, a phenomenon that is genuinely independent of a person's desire cannot depend on this very person's decision to allow it to occur or to prevent it from happening. Obligations and commitments can be all the less agent dependent as they involve other individuals who typically will benefit from having the obligations and commitments fulfilled (Jackendoff, 1999; Jacques, 2001). The self-constitution of the agent as a logical subject of obligation that seems to underlie the expressions such as "I commit, I promise, or I have to pay for my beer" is a superficial, misleading grammatical feature, leading to the oxymoron of a "private obligation."[5] Obligations and commitments comprise in their conditions of satisfaction the actual or virtual presence of beneficiaries counting on their realization, just like any speech act

5. Of course, this point relates back to the well-known argument of "private language" (Wittgenstein 1953/1981; Bouveresse 1976). According to Wittgenstein, the

comprises in its conditions of satisfaction the *recognition* of its validity by its addressees.[6]

This "others-oriented" dimension of desire-independent reasons does not apparently cross Searle's mind, which sweeps away decades of social psychology, sociology, and for once commonsense, by claiming, "Why should I, as a conscious self, care in the least about what other people think I am bound or obligated to do?" (Searle 2001b, 178). Admittedly, we could imagine the unlikely. Phenomenologically, some self-centered individuals could remain totally indifferent to other people's opinion. In accordance with Searle's methodological solipsism, those egocentric persons would be focused only on the phenomenological, experiential counterpart of their obligations and commitments. But phenomenological and logical matters must not be mixed up. Logically, methodological solipsism cannot account for the constitutive relational dimension of obligations and commitments. Obligations and commitments indeed cannot be cut off from the outside world without losing what they structurally are: either a dyadic relationship involving "the second person" of the beneficiary (I made the commitment to go to *your* party) or a triadic relationship involving both another person (the waiter) and an institutional economical rule (*if* I order a beer, *then* I have to pay for it; Descombes, 1996). In fact, those structural relationships, the complementary roles they govern (the roles of customer and waiter), and the rights and obligations they imply (the right of drinking the beer entails the duty to pay for it) explain accurately why commitments and obligations do not have the same direction of fit as intentions and desires. Actually, well beyond

---

meaning cannot be private because an individual could not verify by himself the "sameness" of the terms he would use to refer to his own sensations. Actually, the idiosyncratic convention he would establish with himself would have no criteria of verification apart from his own memory, necessarily prone to retrospective illusions. Only the existence of public, objective criteria in force in a given community is able to check the good usage of a rule and the correct, relevant application of a meaning. This argument, quite convincing for the language of sensations, is still more convincing as regards obligations and commitments, which involve by definition other people.

6. Actually, unless one oddly claims, with Searle (1991), that others' understanding is not an expected effect of one's own utterance or that one can perfectly speak without caring about the reaction of one's interlocutor, any speech act appears to be, by definition, perlocutionary. A speech act aims at having a certain effect on the interlocutor, namely, an effect of recognition. For an interesting discussion of Searle's theory of speech acts, see Habermas (1992/1994). That is also the point that Smith (2003) emphasizes by recalling the additional deontic condition that Reinach adds on the performance of the social act of promising: promises must not be only heard but also accepted by the one to whom it is addressed.

the intersubjective, contractual component of mutual agreements and promises, commitments and obligations are embedded in a holistic web of structural dependences that imposes upon agents the way they are supposed to act.

If desires and intentions can be conceptualized as being "in an individual mind," obligations and commitments mean for the person to be in a certain objective social situation. As Jackendoff (1999, 78) put it, the description "Bill's belief and desire are in his mind" is a tautology, whereas "Bill's obligation is in his mind" attributes to Bill some kind of delusion about his social relations. Just like "truth is a reason for anybody to believe" (Searle 2001b, 137), commitments and obligations are reasons for *anybody* to behave. In the same way that beliefs are commitments to truth and respond to what ought to be believed, practical commitments respond, I claim, to what *ought to be done*. At the exact opposite of desires and intentions, commitments and obligations thus have the same mind-to-world direction of fit within practical reasoning as beliefs have within theoretical reasoning: they have to adjust to the impersonal, objective features that characterize the external reality. After I made the commitment to go to Chicago, "going to Chicago" is a future state of the world to which *I have to* conform: I must adjust my behavior and make sure that I act in the way I ought to act by effectively going to Chicago. The objective commitment to go to Chicago is the antecedent condition that fosters, if not causes, my subsequent action. From this perspective, desire-independent reasons such as obligations and commitments have an "upward" world-to-mind direction of causation that requires less a theory of propositional attitudes than an ontology of normative facts. This causal directionality indeed meets the main criterion that permits us to select the potential candidates for an ontological status: the causal role peculiar to "basic metaphysical categories" (Laurence & Macdonald, 1998). Of course, this argument raises the puzzling question of the place of those strange "causal" entities in the ontological furniture of the world. To explore the ontological status of desire-independent reasons, we have to examine further the specific mode of existence of normative facts. That is the point of our second episode.

IIb.2. Second Episode: The Strange Objectivity of Desire-Independent Reasons

To ask what kind of objectivity desire-independent reasons have, let us first put aside the strange contradiction we find in Searle's the-

ory, in which obligations and commitments are at the same time "neither facts nor intentional states" (Searle 2001b, 103) and "epistemically objective facts in the world" (Searle 2001b, 119). Let us also forget the fact that, contrary to Searle's claim, the subject cannot himself create a reason for acting independently of his desires when the reason in question is described as objective. And let us dwell on the only genuine desire-independent episode of Searle's model, which is about the status of fact peculiar to obligations and commitments.

While granting a factual status to obligations and commitments, Searle does not specify the essential property of those facts, that is, their *normativity* or their capacity to obligate. Thanks to this remarkable omission, normativity can be seen as external to the facts in question: that is the agent who is in charge of turning or not turning this mysterious *is* into an effective *ought*. Yet, pace Searle, a real *ought* is not grafted by agents onto the *is* of a state of affairs, say the fact to make a promise. Actually, the promise is not any fact; it is a fact that can be described either *descriptively* (e.g., my promise to help him has been made) or *prescriptively* (e.g., my promise to help him entails I ought to help him; von Wright 1995). This double status is constitutive of normative facts, whose prescriptive power cannot be accounted for by the interpretative work of individuals or the subjective feeling of being bound. A normative fact does not become normative because a particular agent deigns to recognize it as an *ought to*. A normative fact, I think, is by definition intrinsically normative: it logically comprises both an *is* and an *ought*, as revealed in the conditional format "if promise *p*, then obligation *q*." Instead of being separated by the artificial gap that Searle assumes in order to leave the decisive intervention to the self, the antecedent and the consequent are two sides of the same fact (Petit 1995). In other words, the obligation to pay for a beer is not the correlate of the agent's subjective act of recognition; it is the logical correlate of the very act of ordering a beer.

Given the above, ordering a beer and creating the obligation to pay are not two different actions, brought together by the intentional power of the self, but two parts of one single action, governed by the rule of commercial exchange. Far from following the satisfaction of a prior desire (to drink the beer) in an ulterior sequence made artificially autonomous, the desire-independent commitment is part of the unique whole action "to order a beer in a bar." Unfortunately, Searle cannot account for this holistic dimension of action because he imposes on action the same logical treatment as Zeno used to impose on space and movement.

Let us recall that, for Zeno, a fired arrow can never reach its target because, in order to do that, it needs first to cover the half of the required distance and the half of the half distance and so forth, indefinitely. The structure of this philosophical puzzle, which shows how logical reasoning can sometimes contradict the most elementary commonsense, is found in the way Searle conceives of free will. Actually, to show the inalienability of free will at each moment of our life, Searle splits up whole, meaningful actions into short, successive episodes whose single performance seems to require the full exercise of free will. For instance, the same course of action "to get a beer in exchange of money" is split up into artificial sequences that transform a rule-governed behavior and an institutional obligation into a free will matter. The application of this new version of the Zeno paradox to an action such as "to obey the order of his military superior, who commands his soldiers to march on the enemy's capital" leads to a strange statement. If we follow Searle's reasoning, the soldiers choose, instead of staying motionless, to take a step forward and another step forward and so forth, which implies they do march on the enemy's capital of their own free will.

Of course, such a conclusion is just as counterintuitive as it is ideological. In fact, the soldiers' micro-actions take on their meaning only within the frame of the whole action "to obey their superior." It shows how the determination of what counts as an action or not matters in deciding whether individuals have will free will and also how free will can be explained. If one adopts a micro-scale, focused on meaningless, derisory actions, such as taking a step forward or moving one's finger, one increases the chance to define free will in Searle's way. But the price to pay for this result is the same as the logical immobility of Zeno's arrow: both conclusions are a mere construction of the philosopher's mind. On the other hand, if one selects a larger scale that individuates the action, not its fragments, as a significant whole, one sees those so-called free micro-actions as the steps necessary to perform the global action that makes sense of them. Now, the reasons for performing such global actions, whether getting involved in an economical exchange (such as the beer example) or owing obedience to superiors (like the military case), are clearly not the making of any self. Those reasons are duly provided by economic and military institutions.

Yet Searle denies such a conclusion. It is not, he says, the endorsement of institutions that is the essential source of obligations but the self (Searle 2001b, 198). According to him, institutions provide only the vehicle for the intentional, voluntary creation of obligations, com-

mitments, and duties. They are just instrumental tools that enable individuals to express private thoughts and act in a purposeful way. "The institution of promising, and other such institutions, make it possible for rational agents acting freely to publicly bind their will in the future by acts performed in the present" (Searle 2002, 5). In this respect, Searle's intentionalist conception of language is a good give-away. Like any institution, language is the "syntactic format" allowing to code the self's intentions of meaning and action, a superficial covering that is unable to generate semantic contents or impersonal obligations unless the users freely give their assent to them. "Unless I think English is somehow a good thing, I am under no obligations when I speak it" (Searle 2001b, 199). The problem is that such an instrumentalist perspective not only lapses into the solipsist view of institutions as facultative options. It also rules out of its framework all the reasons for acting that are grounded in a community of institutions, rules, and practices. Those reasons, traditionally called "external reasons," provoke an "adequate reaction," a "correct answer" from the competent members of a community. For instance, to give an answer to a question, to obey to one's superior order, to brake at the sight of a red light, or to keep one's promises are different kinds of adequate reactions to external reasons for acting—in this case, the question, the order, the red light, and the promise (von Wright 1983). In those examples, agents do not use the institutions around them to satisfy their desires; they directly respond to social settings.

Obviously, reactive actions thus triggered or fostered by institutional rules are not free in Searle's sense. To him, desire-independent reasons that are not the result of voluntary exchanges but depend on external conditions must be kept out of the issue of free will. As Zaibert (2003) put it, Searle's insistence on voluntariness is not suited to nonpromissory commitments and obligations whose binding force, although deriving from an institutional source, does override the commitments that might arise from the deliberate act of promising. "For example, I have never promised that I will not intentionally and unjustifiably injure a fellow human being, yet I am wholeheartedly committed to not doing this. This is an enforceable commitment, both legally and morally" (Zaibert 2003, 73). By modeling any kind of desire-independent reasons on the self-created obligations emerging from voluntary speech acts, Searle's theory is not only unable to account for enforceable commitments. It also leads to consider the agent submitted to those impersonal, nonpromissory commitments as being irrational or alienated. Like the slave who never exercises

any freedom in creating a reason for himself to act (Searle 2001b, 210), the agent who does not intend to place himself under an obligation but is passively subjected to it is neither free nor rational. Such reasoning thus results in assuming, as Castoriadis (1988) put it, that 95% of human actions are not rational but traditional. The range of actions in which genuine free will is exercised—and hence Searle's inquiry—is indeed dramatically reduced to a derisory 5% of human actions, namely, subjectively intended micro-actions (i.e., moving one's finger, scratching one's head, or going swimming rather than biking). To avoid this disastrous consequence, it is necessary to admit that desire-independent reasons, although generated by institutions, are able to induce rational actions. In fact, institutional reasons, though following the mind-to-world direction of causation that we spoke of above, exert only a conditional causation over individuals.[7] Agents keep at least the virtual possibility of evaluating the meaning of their actions and envisaging alternative options, possibility which is a central criterion for rationality (Winch 1958).

Given the existence of objective, institutional reasons, it remains to be seen how normative facts can act upon individual minds. After all, to the extent that normative facts are both desire independent and not strictly causal, they require from the agent the capacity to make their normativity effectual. Now, in Searle's instrumentalist model of institutions, it is very difficult to understand why the self would freely force itself into some course of action by recognizing the alleged neutral *is* of desire-independent reasons as an effective *ought*. In my sense, the answer lies in an internal version of institutions, focused on the "background" of actions and on the "causal sensitivity to institutions" that Searle himself, who denies its relevance to the issue of rationality and free will, speaks of elsewhere (Searle 1995, 142).

IIb.3. Third Episode: The Binding Recognition
of Desire-Independent Reasons

Within Searle's framework, the transformation of what *can* and *may be* a reason for action into an internal, effective motivation for act-

7. Conditional causation, as Peirce put it, is not about the positive universe of empirical occurrences: it is about properties, status, and types of actions, which virtually determine the range of future behaviors, not the behaviors themselves (Hulswit 2002). The impersonal entitlements and obligations constitutive of a given status, for instance, husband, rule a whole range of behaviors without causing the effective behaviors of real people, who can always deny or disclaim those entitlements and obligations.

ing is quite mysterious. After all, "why should the commitments that I voluntarily and intentionally undertake be binding on me?" (Zaibert 2003). Searle's unique solution to this issue is the vague notion of "recognitional rationality," that is, a process that is supposed to translate desire-independent reasons into internal motivators. But what about the dynamics of this recognition? To me, the only way to solve this puzzle is to resort to the preintentional "background," composed of practical abilities, extralinguistic expectations, and tacit knowledge that shape intentional states (Searle 1983, 1995, 1998). But unfortunately for the coherence of Searle's whole theory, the background miraculously vanishes from his argumentation as soon as the issues of free will and practical deliberation are concerned. While admitting the background can strongly limit the range of possibilities getting to the level of consciousness, Searle (2001b, 25) claims that this limitation is not a logical or theoretical issue but only a practical issue. Yet, in my view, the omnipresence of the background contradicts both practically and logically Searle's strange assumption according to which "we are presented with an indefinite, indeed strictly speaking infinite, range of choices" (Searle 2001b, 233).

In fact, I think that the role of the background is anything but a contingent, practical issue. Searle, who has emphasized in the past the logical importance of the background, should have seen this point. Actually, in his first works, the existence of the background is postulated to solve the logical problem of the infinite regress of representational stages, in which the indeterminate interpretation of one stage must be completed by a further indeterminate interpretation ad infinitum. Given that the management of an infinite series of successive interpretations is pragmatically, evolutionary, and cognitively impossible, organisms must logically have at their disposal implicit resources that complete automatically the informational inputs (Searle 1983). For Searle, this set of resources, called the background, restrains the range of possible interpretations, imposes categorizations on conscious experiences, and shapes new events into understandable phenomena.

The conclusion that everyone—except, curiously, Searle himself—can draw from the preintentional implementation of the background is easy enough: the relationship, if any, that the person can possibly maintain with her own intentional states and hence with her reasons for acting is partially "opaque." Explicit motivations and conscious reasons for acting thus have no practical and logical means to escape

from the rigid conative and cognitive paths structuring conscious mental states in a non- or preintentional way. Moreover, contrary to what Searle (2001b) claims, the logical contribution of the background to the scope of choices and sense of possibilities does make the recognition of desire-independent reasons less enigmatic. This recognition is fostered, I think, by the inclination to acknowledge some facts as worthy of being acted upon. Practical reasoning about the hierarchy of ends, far from starting from scratch, is grounded on a previous set of preferences and dispositions that impose its own scale of priorities on any ulterior decision making. While remaining partially independent of desires, desire-independent reasons take a moral form that makes them, if not desirable, at least intrinsically valuable. As Durkheim put it, moral desirability and obligation are two complementary sides of institutional facts: "obligation or duty express only one of the aspects, and an abstract aspect, of morality. A certain desirability is another feature, no less essential, than the first one" (Durkheim 1924/1996, 50).

Of course, the background contribution to the desirability of certain ends and reasons for action is not idiosyncratic. In my sense, the preintentional set of preferences results in one way or another from the public hierarchization of ends and activities that constitutes, broadly speaking, politics. Actually, politics, defined after Aristotle as an "order of subordination" whose nature is "architectonic," establishes the obligations of individuals toward the city and submits individual activities to public reason (Descombes 1994, 153). Although it would be abusive to turn desire-independent reasons for acting into "reasons of State," such a public, impersonal hierarchy of ends does certainly resonate with the alleged private hierarchy of preferences. As Korsgaard (1996) points out, agents are the "citizens" of "a Kingdom of Ends," a kingdom that is not only moral but also political. The preintentional capacities of the background allow agents to recognize rules and obligations as being good reasons for acting, that is, valuable and justifiable. In other words, desire-independent reasons appear as appropriate candidates for the title of internal motivators because the background is previously sensitive to their appropriateness. This background sensitivity results from the long-term adaptation of individuals to their sociocultural environment, including institutions and the hierarchy of ends these very institutions establish. Institutions, thanks to the prereflexive counterpart of their constitutive norms and obligations, do not need to resort to violence to ensure

their binding force. The prepolitical workings of the background are effective enough to determine desirable reasons for acting, to motivate people to do or not to do certain things, and to make a personal virtue out of institutional necessity.

III. FOR AN ALTERNATIVE MODEL OF AGENCY

By placing Searle's arguments one step after another in proper order, we have seen that his conception is too individualistic to really account for the impersonal, public status of reasons. Reasons for acting, which stem from objective relations of structural dependence and/or common agreement, cannot be dependent on the idiosyncratic goodwill of specific individuals. In a way, Searle's previous thinking backfires on his recent model of free will, devoted to the rescue of the sovereign self. For from my point of view, the canonical form "count-as," which reflects, for Searle (1995), the logical skeleton of institutional facts, applies perfectly well to reasons for acting: to recognize a reason for acting amounts to counting it as a reason. What counts as a reason for acting, far from depending on the self's creativity, is determined by collective practices, institutions, and norms. Reasons for acting thus are not private, mental entities but public, shareable, and generalizable grounds that all can share as participants in a community of practice and language. As a result, the difference of kind that Searle postulates between desire-independent and desire-dependent reasons turns out to be rather a difference of degree. Reasons for acting, in virtue of their public, impersonal dimension, do not have a subjective ontology but what I call a first-person plural ontology (Kaufmann 1999b): like other normative facts, they are irreducible to the mental contents furnishing the narrow, opaque context of a particular mind. Reasons for acting, normative facts, and institutions that frame, supply, and shape individual wills make sense only within the broad context of the *We* of a given community. Even if reasons for acting have been created by human minds, they enjoy high-level properties that cannot be restricted to the mental realm of any solipsist self. In this respect, the analogy between the relationship brain/mind and individual/society that Durkheim (1924/1996) drew many years ago still stands. Just like the high-level mental properties of the mind emerge and are liberated from the low-level configuration of the neurons in the brain, norms and institutions have col-

lective properties that emerge from the interactions between minds and are irreducible to mental properties (Durkheim 1924/1996, 34). The study of psychological processes is indeed as unable to make sense of the action "to order a beer" as the study of the chemical components of money is pointless for the understanding of economical cycles.

If, as I suggest it, only a first-person plural or We-ontology is able to allow for reasons for acting, common meanings, and public practices, phenomena that do respond to a subjective ontology become restricted to one unique domain: the phenomenal domain of sensations and *qualia*, that is, the idiosyncratic properties characterizing *what it is like* to drink a beer or to feel compelled to pay for it. But the *qualia* are far from doing justice to the intentional ability to "get outside the circle of our own ideas" and to be aware of something that is not ourselves (Chisholm 1972, 202). The subjective viewpoint of a transcendental self contemplating its inner realm cannot account for the feature of *aboutness* proper to intentionality. Intentional phenomena indeed comprise by definition distal, external properties, such as the externality of referents, the indexicality of contexts, and the adjustment of interactions. In the absence of those exogenous, shareable, and public properties, individual minds would turn into irrational, ill-adapted selves, unable to make a difference between their own hallucinations and the external reality. To avoid this solipsist tragedy, I will plead for an alternative model of agency after I have reviewed Searle's arguments in favor of the self.

If we follow Searle (2001b, 59), agents have the phenomenological experience that psychological antecedents of their action are not sufficient to determine their action and that alternative future decisions and actions are open to them. Although he admits such an experiential gap could be an illusion, it is an illusion so necessary that we have to presuppose that "the phenomenology corresponds to a reality" (Searle 2001b, 71). In fact, the use of this phenomenological argument is quite surprising for someone who is used to energetically denying the relevance of phenomenology, which is focused on the superficial seeming of the world instead of its real structure (Searle 2001a). But above all, it is well known that the phenomenal account of what it is like to act cannot be turned into the ontological account of what an action really is. As the grammatical approaches, inspired by Wittgenstein, regularly recall it, the conception of the self as a privileged epistemic entity might be nothing more than a superficial fea-

ture of the language game of action (Bouveresse 1976/1987; Ryle 1949/1984; Shanker 1991).[8]

Potentially aware of this difficulty, Searle uses a kind of argumentative hopscotch to skip it. As soon as the kind of argument he is using, in this case phenomenological, becomes dubious, he jumps to another "argumentative square" and turns to a logical argument for help. But this theoretical jump does not prove to be very helpful. Actually, we have seen in the first part of our article that an ontological claim, that is, the self as a substantial entity initiating and carrying out actions, cannot be derived from the self as a formal, logical requirement on rational action. Yet Searle's reasoning goes as follows: from the phenomenological distinction between the events *happening to me* and the actions *I make happen*, and from the logical assumption that each active voluntary process *must* have a self who thinks, he draws the conclusion that such a self really *does exist*. Thanks to this double fallacy, he succeeds in making both the phenomenological feeling of agency and the logical presupposition of actions, that is, responsibility and freedom, dependent on the existence of a conscious, substantial self.

For the people wishing to defend, as I do, a more convincing model of action, what is at stake is to show that neither the phenomenological aspect of agency, namely, the feeling of deciding and acting, nor the presumption of responsibility and freedom require a substantial self. And in fact, there is a theory of action that permits the elimination of any ontological link between agency and a privileged self. This theory, mostly explored by Wittgensteinian philosophers, advocates a teleological rather than a mentalist account of action. Whereas the mentalist model of action holds that a given state of mind determines why one should do X, the teleological model assumes that states of mind are determined by the facts and processes that are suited to achieving one's purpose—say, buying flowers is suited to cheering up someone (Butterfill 2001). Unlike the mentalistic model, the teleological approach avoids artificially breaking down behaviors into two separate and independent actions, one mental and the other one physical. In this view, the intended action is not the physical consequence of a mental plan but the antecedent of a practical reasoning establishing what should be done to perform the action in question

8. In the same way that a sentence such as "the actual king of France is bald" leads us to assume the existence of a king of France, the I-sentences would lead us to assume the existence of an extralinguistic reference, in this case, the self, to the I in position of subject.

(Descombes 1996). Now, the different steps necessary to achieve one's end are too indeterminate to be fixed a priori in the individual's mind: they indeed require successive adjustments to other people's contributions, environmental resources, and contextual constraints. For instance, the agent who intends to converse with his friend or to order a beer in a bar has to deal with the bad mood of his interlocutor or the stress of the waiter to achieve his goal. According to this teleological conception of action, practical rationality is not, as Searle (2001b, 30-31) suggests it, the choice, determined by the self prior to acting, between conflicting or incompatible ends. Practical reasoning obeys, so to speak, a reality principle: it is synonymous with a "trying to get," which implies by definition the functional adaptation to real circumstances (Anscombe 1957; Ricoeur 1977). Agentive power thus lies in the practical capacities for anticipating and evaluating the fallouts of one's action, taking advantage of contingent resources, and determining the most appropriate way to act.

In this framework, the distinction between purposive actions and passive events is no more dependent on the existence of a self "cooking up in its internal kitchen" a decision and using only "the ingredients that it has allowed through the door" (Dennett 2003, 123). Such a distinction emerges from the sense of purposiveness that goes with the practical pursuing of a particular course of action. Phenomenologically, indeed, the feeling of deciding and acting, instead of emerging from an abstract activity of deliberation, loaded with the mysterious "metaphysical weight" of the self, is the counterpart of the act of doing. In the same way, the experience of freedom or non-freedom arises when the agent comes up against external constraints, either other people's wills or institutional forces. Logically also, the teleological perspective avoids creating an ontological illusion, such as Searle's self. Actually, the issues of responsibility and freedom cannot be addressed in the absolute and concerned with a brain isolated in a vat. This is only the performance of a given action in real-life situations that makes those issues relevant, mainly by inevitably raising the questions "of whom this action is" or "to whom is happening" (Descombes 1991).

This claim makes it possible for me to stress an important point. In Searle's account, the issue of free will is framed only on the potential incompatibility between the causal determinism of the physical world and the alleged indeterminacy of ontologically subjective phenomena. Although this framing still prevails in philosophy of mind, it fails to account for the fact that the presumptions of responsibility and

freedom are not formed in a social vacuum. Those presumptions indeed stem from the mutual expectations of accountability and justifiability that actors impose on one another (Baynes 2001). Via the social practice of demanding and giving reasons, individuals are given responsibility over time, so that they can reliably be there to take responsibility and answer when questions of accountability arise (Dennett 2003). The order of explanation here is thus the opposite of Searle's: a person is responsible because people, authorities, and institutions take her to be responsible, which means that a person's being responsible is a function of a stance others take toward her. In other words, the first-person stance adopted by people who claim responsibility for their actions derives from the third-person stance taken by others toward them (Strawson 1959/1990; Kaufmann 1999a). Little by little, individuals appropriate the status of responsible subject ascribed to them by others and adopt the social norms of reasonableness and accountability with regard to themselves. As competent members of their community, they transform the heteroascriptions they are subjected to into "self-ascriptions" (Harré 1983).

The emphasis on freedom and responsibility as creations of an interpersonal design process shifts the issue of free will away from the pursuit of a self whose God-like power would enable it to exempt itself from the causal fabric of the physical world. Once free will has been reframed in terms of "civil" freedom and mutual responsibility, the main problem at stake becomes "first-person authority": this concept, often mentioned in philosophy of action, refers to the fact that individuals, although endorsing hetero-ascriptions and public reasons for acting, enjoy nevertheless an authority about their own thoughts and sensations that is denied to others (Anscombe 1975; Davidson 1984). Even if the description and explanation of a given action necessarily take place in the public space of reasons, the final answer to the question *why* the agent performed such or such action belongs to him (Ricoeur 1977; Kaufmann 1999a). This being so, first-person authority, while respecting the fundamental asymmetry between first- and third-person perspectives, makes sense only in relation to other kinds of authority, whether second-person authority (other people) or collective, impersonal authority (norms, institutions). Unlike selfhood, which obeys a logic of introspection, first-person authority obeys a logic of mutual attribution that emerges from the social practice of holding people as responsible reason givers.

The concept of first-person authority also makes it possible to show that claiming, like Searle (2001b, 178), that "the fact that people

are bound as husbands, wives, citizens, taxpayers . . . say nothing about the first-person point of view" is false. Actually, the attributions and expectations constitutive of institutional roles we are committed to are sooner or later translated into a conception of ourselves as husband, friend, or citizen. Such self-conceptions have obvious effects on the first-person point of view: to endorse the status of husband, friend, or citizen is tantamount to adopting fundamental standards of evaluation and principles of choice as for which actions and reasons for acting are suitable. But contrary to the asocial self, first-person authority is not jeopardized by the ongoing process of appropriating role ascriptions and "reflexively self-referring" that Martin and Sugarman (1999) speak of. First-person authority is a perpetual social achievement that precisely fits people capable of responding to the range of statutory entitlements and obligations, whether official or not.

CONCLUSION

After examining Searle's ontological approach, it seems to me that his reasoning creates, instead of discovers, dualistic gaps. Actually, we have seen first that the causal relations that are supposed to bridge the gap between first-person and third-person phenomena fail to save the alleged monist ontology advocated by Searle's biological naturalism. One way to break this deadlock might be to stop seeking the ontological noncorrespondence between brain and mind. Given the obvious fact that mental states are nothing but physical, Searle's criterion for free will, namely, its independence from the brain processes that embody it, sounds wrongly dualistic. In this respect, the focus on the sensory, functional mappings that define the brain/mind as a dynamic whole, able to direct living organisms in their environment, seems more promising (Proust 1997; Damasio 1999). Within this functionalist framework, the issue of free will—that is, how can free will survive the causal, blind forces of physical facts—yields to the issue of conscious will: does the fact that certain states are conscious give them a causal, decisive power over unconscious states? In other words, does the conscious access to certain mental states exert a kind of feedback causation on the mostly unconscious action processing? According to Libet (1999), this kind of feedback does exist. Even if, as he has shown in previous works, the initiation of a voluntary act begins in the brain well before the person consciously knows she

wants to act, the conscious function has enough time to affect the final outcome of the volitional process. In Libet's account, conscious will, although it does not initiate a voluntary act, can thus veto the process and control whether the act takes place. From my point of view, such a "veto-consciousness" has the great advantage of accounting for agency without requiring a sovereign self ruling a free market of transparent reasons.

In the second part of our article, we have seen that Searle's ontology of reasons for acting feeds on the solipsist gap created between the inner, subjective realm and the outer, social world. The problem is that the surprising indifference of individuals to other people and institutions is not only questionable for the philosophers who have what Searle (2001b, 182) calls an "unhealthy obsession with something called ethics and morality." It is problematic for any theory of action for it makes the self both irresponsible and asocial. Actually, Searle's solipsist self can place its private intentions in opposition to other people's interpretation and disclaim the public consequences of its own actions. Whereas the slave that Searle speaks of is not accountable for his actions because he is forced to act, the self cannot be held accountable for its actions whenever *it does not want to*. Counter to this view, I have argued that social, real agents cannot escape from responsibility and accountability. The command of the course of any meaningful action does require the identification of good reasons for acting and the anticipation of the potential justifiability of one's behaviors. Far from the exclusive property of a solipsist self, the capacity to determine the sense, reasonableness, and desirability of one's action comes within the competence that any member of the community must master: the competence to endorse the We-stance or, as Mead (1962/1970) put it, the impersonal viewpoint of the "generalized other." Instead of making the I-stance disappear, the We-stance is its logical condition of possibility. First-person authority indeed emerges little by little from the objective reality of norms, obligations, and roles that enable individuals to shape their identity through mediated self-references. Reconceptualized under the aegis of first-person authority, agency, freedom, and responsibility do not require the maintenance of a field of gaps. They appear as fundamentally social phenomena whose ontology is neither the third-person ontology of the matter nor the first-person ontology of phenomenal consciousness. Rather, it is the first-person plural ontology specific to the impersonal properties of normative and institutional facts, including

reasons for acting, that make sense of agency, freedom, and responsibility.

REFERENCES

Anscombe, E. 1957. *Intention*. Oxford, UK: Oxford University Press.

———. 1975. The first person. In *Mind and language: Wolfson College Lectures 1974*, edited by S. Guttenplan. Oxford, UK: Oxford University Press.

Baynes, K. 2001. Practical reason, the space of reasons and public reason. In *Pluralism and the pragmatic turn: The transformation of critical theory*, edited by W. Rehg and J. Bohman. Cambridge, MA: MIT Press.

Bouveresse, J. 1976/1987. *Le mythe de l'intériorité*. Paris: Minuit.

Butterfill, S. 2001. Two kinds of purposive action. *European Journal of Philosophy* 9 (2): 141-65.

Castoriadis, C. 1988. Individu, société, rationalité, histoire. *Esprit* 2:89-113.

Chisholm, R. 1972. Intentionality. In *The Encyclopedia of Philosophia*. New York: Macmillan

Clément, F., and L. Kaufmann. 1996. Esquisse d'une ontologie des faits sociaux: La posologie proposée par John Searle. *Réseaux* 76:123-61.

Collin, F. 1997. *Social reality*. London: Routledge.

Corcoran, K. 2001. The trouble with Searle's biological naturalism. *Erkenntnis* 55:307-24.

Damasio, A. R. 1999. *The feeling of what happens: Body and emotion in the making of consciousness*. London: William Heinneman, Random House.

Davidson, D. 1984. First person authority. *Dialectica* 38 (2-3): 101-11.

Dennett, D. 2003. *Freedom evolves*. New York: Viking Press.

Descombes, V. 1991. Le pouvoir d'être soi. *Critique* 529-530:545-76.

———. 1994. Philosophie du jugement politique. *La Pensée Politique* 2:131-57.

———. 1996. *Les institutions du sens*. Paris: Minuit.

Durkheim, E. 1924/1996. *Sociologie et philosophie*. Paris: PUF.

Esfeld, M. 1998. Holism and analytic philosophy. *Mind* 107:365-80.

Habermas, J. 1992/1994. *Postmetaphysical thinking: Philosophical essays*. Cambridge, MA: MIT Press.

Harré, R. 1983. *Personal being*. Cambridge, MA: Harvard University Press.

Hulswit, M. 2002. *From cause to causation: A Peircean perspective*. Dordrecht, the Netherlands: Kluwer Academic.

Jackendoff, R. 1999. The natural logic of rights and obligations. In *Language, logic, and concepts: Essays in memory of John Macnamara*, edited by R. Jackendoff and P. Bloom. Cambridge, MA: MIT Press.

Jacques, F. 2001. Remarques sur la promesse et le pardon. La théorie des actes de langage à l'épreuve de l'éthique. *Revue Internationale de Philosophie* 2 (217): 227-42.

Kaufmann, L. 1999a. Esprit, es-tu là? Le sociologue et l'autorité de la première personne. *Information sur les Sciences Sociales/Social Science Information* 38 (2): 203-48.

———. 1999b. Les faits sociaux sont-ils des choses? *Critique* 630:890-911.

Korsgaard, C. M. 1996. *The sources of normativity*. Cambridge: Cambridge University Press.

Laurence, S., and C. Macdonald. 1998. Introduction: Metaphysics and ontology. In *Contemporary readings in the foundations of metaphysics*, edited by S. Laurence and C. Macdonald. Oxford, UK: Blackwell.

Libet, B. 1999. Do we have free will?. *Journal of Consciousness Studies* 6 (8-9): 47-57.

MacIntyre, A. 1986. The intelligibility of action. In *Rationality, relativism and the human sciences*, edited by J. Margolis, M. Krausz, and R. M. Burian. Dordrecht, the Netherlands: Martinus Nijhoff.

Martin, J., and J. Sugarman. 1999. Is the self a kind of understanding? *Journal for the Theory of Social Behaviour* 31 (1): 103-14.

Mead, G. H. 1962/1970. *Mind, self, and society*. Chicago: University of Chicago Press.

Mills, W. C. 1940. Situated actions and vocabularies of motive. *American Sociological Review* 5:904-13.

Nida-Rümelin, M. 2002. Causal reduction, ontological reduction, and first-person ontology. In *Speech acts, mind, and social reality: Discussions with John R. Searle*, edited by G. Grewendorf and G. Meggle. Dordrecht, the Netherlands: Kluwer Academic.

Olafson, F. A. 1994. Brain dualism. *Inquiry* 37:253-65.

Penrose, R. 1994. *Shadows of the mind: A search for the missing science of consciousness*. New York: Oxford University Press.

Petit, J.-L. 1995. Introduction. In *La querelle des normes*, edited by J.-L. Petit. Caen, France: Presses Universitaires de Caen.

Proust, J. 1997. *Comment l'esprit vient aux bêtes: Essai sur la représentation*. Paris: Gallimard.

Putnam, H. 1975. The meaning of meaning In *Mind, Language and Reality. Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press: 215-271

Récanati, F. 1984. Pour la philosophie analytique. *Critique* 444:362-83.

Ricoeur, P. 1977. Le discours de l'action. In *La sémantique de l'action*, edited by D. Tiffeneau. Paris: CNRS.

Ryle, G. 1949/1984. *The concept of mind*. Chicago: University of Chicago Press.

Searle, J. R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

———. 1983. *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

———. 1990. Collective intentions and actions. In *Intentions in communication*, edited by P. Cohen, J. Morgan, and M. E. Pollack. Cambridge, MA: MIT Press.

———. 1991. Reply to D. M. Armstrong. In *John Searle and his critics*, edited by E. Lepore and R. Van Gulick. Cambridge: Basil Blackwell.

———. 1992. *The rediscovery of the mind*. Cambridge, MA: MIT Press.

———. 1995. *The construction of social reality*. New York: Free Press.

———. 1998. *Mind, language, and society: Philosophy in the real world*. New York: Basic Books.

———. 2000. Consciousness. *Annual Review of Neurosciences* 23:557-78.

———. 2001a. Neither phenomenological description nor rational reconstruction: Reply to Dreyfus. *Revue Internationale de Philosophie* 2 (217): 277-84.

———. 2001b. *Rationality in action*. Cambridge, MA: MIT Press.

———. 2002. Speech acts, mind, and social reality. In *Speech acts, mind, and social reality: Discussions with John R. Searle*, edited by G. Grewendorf and G. Meggle. Dordrecht, the Netherlands: Kluwer Academic.

Shanker, S. 1991. The enduring relevance of Wittgenstein's remarks on intentions. In *Investigating psychology: Sciences of mind after Wittgenstein*, edited by J. Hyman. London: Routledge.

Smith, B. 2003. John Searle: From speech acts to social reality. In *John Searle*, edited by B. Smith. Cambridge: Cambridge University Press.

Strawson, P. 1959/1990. *Individuals: An essay in descriptive metaphysics*. London: Routledge.

von Wright, G. H. 1983. *Practical reason*. New York: Cornell University Press.

———. 1995. Y a-t-il une logique des normes? In *La querelle des normes*, edited by J.-L. Petit. Caen, France: Presses Universitaires de Caen.

Winch, P. 1958. *The idea of a social science and its relation to philosophy*. Atlantic Highlands, NJ: Humanities Press.

Wittgenstein, L. 1953/1981. *Philosophical investigations*. Oxford, UK: Blackwell.

Zaibert, L. 2003. Intentions, promises, and obligations. In *John Searle*, edited by B. Smith. Cambridge: Cambridge University Press.

Ziff, P. 1960. *Semantic analysis*. New York: Cornell University Press.

After several years of research in Paris (EHESS) and in the United States (Berkeley, Michigan, and Harvard), Laurence Kaufmann now teaches at the University of Lausanne, Switzerland. Her publications are concerned primarily with the ontology of social facts, from a historical and sociological point of view (how institutions emerge and change over time) as well as from a philosophical and psychological perspective (what abilities human beings need to build up and maintain the strange "entities" that institutions are). She recently edited, with J. Guilhaumou, a collection of essays titled L'invention de la Société: Nominalisme Politique et Science Sociale au 18ème Siècle *(EHESS, 2003).*