Christine Clavien

Kitcher's Revolutionary Reasoning Inversion in Ethics*

Abstract: This paper examines three specific issues raised by *The Ethical Project.* First, I discuss the varieties of altruism and spell out the differences between the definitions proposed by Kitcher and the ways altruism is usually conceived in biology, philosophy, psychology, and economics literature. Second, with the example of Kitcher's account, I take a critical look at evolutionary stories of the emergence of human ethical practices. Third, I point to the revolutionary implications of the Darwinian methodology when it is thoughtfully applied to ethics.

1. Introduction

> "Ethics emerges as a human phenomenon, permanently unfinished.
> We, collectively, made it up, and have developed, refined, and dis-
> torted it, generation by generation. Ethics should be understood as
> a project—the ethical project—in which we have been engaged for
> most of our history as a species." (*EP*, 2)

As revealed by this introductory sentence, Philip Kitcher is one of the few philosophers who take the implications of the Darwinian revolution for moral theory at face value. He does not simply fit his theory to a preconception of *human nature*, but adapts his thoughts to *human evolution*. The latter is a rad- ically different approach since evolution is an on-going process and thus calls for scepticism regarding fixed views on what human nature is supposed to be.

At least three key messages are contained in *The Ethical Project.* First, ethics is tightly linked to altruism in a way that Kitcher aims to clarify. Second, the roots of ethics are to be found in our evolutionary past and some of its crucial elements can be observed in our near cousins, the great apes. Third, the descriptive analysis of human ethical practices has important implications for other areas of ethics. In what follows, I propose to take up these three topics. First, regarding Kitcher's treatment of altruism, I'll make an attempt to clarify some elements that might be misunderstood by his readers. Second, I'll voice

some reservations about Kitcher's evolutionary account of ethics. Third, I'll highlight one truly revolutionary aspect of Kitcher's 'ethical project': a reasoning inversion that has important implications for metaethics, normative ethics and practical ethics.

2. Kitcher's Account of Altruism Contrasted

One essential building block of Kitcher's 'ethical project' is altruism. Altruism is an ambiguous term because it is used in many disciplines (i.e. biology, philosophy, economics) and given different meanings depending on the theoretical contexts in which it is embedded. This proliferation of meanings is particularly problematic for researchers interested in a naturalistic account of human ethical practices because all these different concepts of altruism may be integrated into such an account, each playing a different explanatory role.

In order to avoid conceptual confusion, Kitcher begins his book with useful distinctions between three ways of understanding altruism, which he labels *biological altruism*, *behavioural altruism*, and *psychological altruism*. These three definitions help him to pinpoint the exact levels at which experimental data and scientific theories can be recruited in his 'ethical project'. As such they are necessary theoretical working tools. However, we should resist the temptation to assume that these definitions correspond closely to how altruism is conceived in the different fields that have traditionally dealt with altruism.

Firstly, Kitcher's *psychological altruism* is far richer and more differentiated than the characterization typically used in psychological and philosophical literature. Philosophers usually frame the debate around psychological altruism within the framework of a commonsense 'belief-desire psychology'. An action is produced by a single causal chain, starting with a primary desire which determines whether the action earns the altruistic label: only other-directed primary desires confer altruistic status on the whole action (Stich et al. 2010; Clavien 2012). This picture of human motivation ignores many relevant aspects of other-regarding motives, such as the role of caring and empathic emotions, or the fact that one can be more or less disposed to help others depending on various dimensions (e.g. physical distance, level of personal investment). In contrast, Kitcher integrates these elements into his account of psychological altruism. For him, altruism is a capacity to align one's motivational states of mind more closely with others' wants and needs: "When you come to see that what you do will affect other people, the wants you have, the emotions you feel, the intentions you form, change from what they would have been in the absence of that recognition." (*EP*, 20) According to this view, two versions of psychological altruism are possible: emotional altruism (discussed on pages 25–30), which might be found in different animal species, and what one might call desire altruism (discussed on pages 21–23), a cognitively more demanding form that is about the modification of desires. Kitcher also considers altruism a multidimensional phenomenon that depends on factors such as the 'intensity' of one's response to others, the 'range' of those towards whom we are prepared to respond altruistically, the 'scope' of

contexts in which one is disposed to respond, the 'discernment' in appreciating the consequences for others, and the 'empathetic skill' in identifying the desires others have or the predicaments in which they find themselves (discussed on pages 31–34). By spelling out these various aspects of psychological altruism, Kitcher makes a break with philosophical tradition, further enriching his analysis with inputs from empirical studies on humans and social animals.

Secondly, Kitcher's characterization of *biological altruism* does not correspond to the use of this word among the most prominent biologists working on altruism.[1] For Kitcher, "an organism A is biologically altruistic toward a beneficiary B just in case A acts in ways that decrease its own reproductive success and increase the reproductive success of B" (*EP*, 18). This definition does not impose a decrease in reproductive success over the individual's whole lifetime. Thus, a disposition to give up something in situations where the agent is likely to receive a greater gain in exchange in subsequent encounters earns the altruism label. One famous example is 'reciprocal altruism', a form of cooperative behaviour that is beneficial in the long-term described by Robert Trivers (1971). It is important to note, however, that this form of behaviour does not exemplify altruism in the classical sense, defined by William Hamilton when he developed his kin selection theory (1964), and shared by most contemporary biologists working on altruism (Frank 1998; Grafen 1985; Lehmann/Keller 2006; West et al. 2007; Lehmann et al. 2006). For these researchers, altruism refers to a behaviour that *permanently* decreases the actor's reproductive success and increases other organisms reproductive success; that is, the number of offspring is calculated at the end of the organisms' lives (Clavien/Chapuisat 2012). In this much more restrictive sense, helping behaviour that pays off in the long run cannot be considered altruistic. Conceptual clarification is important here because I fear that Kitcher's definition of biological altruism serves as vector for perpetuating a misunderstanding that is frequent in the humanities and social sciences. Scholars are often unaware that Trivers (1971) and Hamilton (1964) did not discuss the same biological phenomenon in their seminal papers; having heard of Trivers' theory (and having often misunderstood Hamilton's, which requires basic knowledge of population genetics), these scholars infer that biologists deny the possibility of any form of individually costly helping behaviour.[2] I hope to have made it clear that this conclusion is mistaken.

Thirdly, how well Kitcher's *behavioural altruism* matches the way economics conceives altruism is a more delicate question, since economists themselves do not seem to be clear on this point (more on this topic in Clavien/Chapuisat forthcoming). According to Kitcher (*EP*, 23):

> "Behavioral altruists are people who look like psychological altruists. That is, they perform the actions people with psychologically altruistic desires would have been led to perform. In ascribing behavioral altruism, however, we do not suppose any particular psychological

---

[1] Unfortunately, Kitcher is ambiguous on this point (see *EP*, 18 and 49).

[2] Note that Kitcher himself does not make this error, despite the fact that he groups the two distinct phenomena under the same broad definition of altruism.

> explanation of the actions. Perhaps they are indeed the products of
> psychologically altruistic desires, or perhaps the actions are produced
> by [...] a desire for status, or for feeling oneself in accordance with
> some socially approved pattern of conduct, or even a self-interested
> calculation."

This description seems to refer to traditional neoclassical theory in economics which ascribes *utility functions* to humans as a way of making predictions about their behaviour: it is assumed that people behave so as to maximize their utility function. A utility function is an axiomatic way to represent people's long-standing and hierarchically ranked *revealed preferences*. Revealed preferences capture stable patterns of choices. For instance, if subjects reliably prefer more monetary units over less monetary units, they reveal a preference for increasing monetary income. Thus at least part of their utility is to increase monetary income.

However, this way of modeling human choices does not make strong assumptions about the *subjective preferences* underlying revealed preferences. For instance, by choosing to increase their income, subjects might desire more money, or they might be motivated by a desire to follow a social rule; the theory does not discriminate between these (or other) possibilities. As the economist Ken Binmore points out,

> "All that is necessary for the theory to apply is that people behave
> consistently. It can then be shown that they necessarily behave as
> though maximizing the expected value of something, whether they
> are intending to or not. Whatever this abstract something may be
> in a particular context, we call it utility." (Binmore 2005, 65)

Thus, Kitcher's behavioural altruists show other-directed revealed preferences, although it is not clear which intimate motives lie behind these preferences (long term self-interested calculation, genuine desire to help others, etc.).

In recent years, however, things have become less clear with the emergence of behavioural and experimental economics. Scholars in these research fields[3] still contribute to the elaboration of long-standing and hierarchically ranked preferences as indicators for actual decision-making. But, in addition, they also tend to define these preferences in a more realistic way; these preferences are more or less explicitly assumed to represent meaningfully what happens in humans' minds while they make their choices. Consequently, these scholars propose refining the utility function by adding more differentiated preferences than the standard preference for monetary gain used in most economics models. For example, if other-directed behaviour is observed in laboratory money games, behavioural and experimental economists incorporate a taste for fairness in the utility function they attribute to humans (Fehr/Gächter 2002; 2005); similarly, if it is observed that subjects who decide to apply costly punishment against free-riding also show activation of a brain area known to correlate with expectation

---

[3] Note that it is precisely to these research fields that Kitcher refers while discussing 'behavioral altruism' (*EP*, 45).

of reward, a taste for revenge or for warm glow of being a caretaker of justice is added in the utility function (de Quervain et al. 2004). In principle, there is no limit to how many preferences can be introduced (and hierarchized) in a utility function. The more complex and fine-tuned it becomes, the more likely it correctly describes the real motivational mechanisms underlying decision making.[4] We can see here that there is a very thin line demarcating *observed preferences* (as simple predictors for action) from *subjective preferences* (as correct descriptions of individuals' private motives). In the literature, it is not always easy to identify whether this line has been crossed or not (Calvien/Chapuisat forthcoming), especially since the increasing use of novel neuroimaging technologies that are meant to provide a glimpse into people's private minds (Clavien/Klein 2010).

To illustrate this ambiguity, it is worth comparing how Kitcher and economics-minded authors formalize altruism. Kitcher proposes a simplified formalization of psychological altruism (in contrast to behavioural altruism). His idea is to represent people's subjective preferences using "real numbers that correspond to how much they value a given outcome". As he explains (*EP*, 24):

> "The numbers assigned in my social desires would be given by the simple equation:
>
> $$v_{Soc} = w_{Ego} \ v_{Sol} + w_{Alt} \ v_{Ben}$$
>
> where v_{Soc} measures my social desires, v_{Sol} my solitary desires, v_{Ben} the measurements of desire I attribute to the beneficiary (you), and w_{Ego} and w_{Alt} the weights given to my solitary desires and my attributions of desire values to you (so that $w_{Ego} + w_{Alt} = 1$)."

Here, w_{Alt} can be conceived as the degree of altruism. Interestingly, some authors from economics provide strikingly similar formalizations for altruistic behaviour. They build utility functions that contain social and other-regarding incentives and explicitly oppose them to self-directed (including so called 'warm-glow') utility functions. For example, Mayr et al. (2009) construct a function as the sum of the actor's welfare and the welfare for the poor weighted by an other-regarding factor. This idea can be formalized with the following equation:

$$u_1(\mathrm{x}) = w_1(\mathrm{x}) + \beta \ w_2(\mathrm{x})$$

where $u_1(\mathrm{x})$ describes the actor's utility (that is, how much she values the outcome of the action x), $w_1(\mathrm{x})$ the actor's welfare if action x is performed, $w_2(\mathrm{x})$ the welfare for the poor if action x is performed, and β how much the actor cares for the welfare of others (if $\beta = 0$, she has no other-regarding preference). Here, β can also be conceived as the degree of altruism.

Similarly, in an analysis of the evolutionary stability of other-regarding preferences in two person interactions, Akçay et al. (2009) propose representing 'how much an individual 'likes' a given outcome' ($u_1(\mathrm{x})$) as the product of her own

---

[4] However, one drawback that prevents many economists from developing complex utility functions is the lack of predictive power of functions composed of multiple variables.

payoff and her partner's payoff weighted by an other-regarding factor. This idea can be formalized with:

$$u_1(\mathrm{x}) = w_1(\mathrm{x})\ w_2(\mathrm{x})^{\beta}$$

where $w_1(\mathrm{x})$ is the actor's payoff if action x is performed, $w_2(\mathrm{x})$ is the second player's payoff if action x is performed, and β measures the 'degree of the other-regarding preference' of the actor for her partner (if $\beta = 0$, she has no other-regarding preference). Here again, β seems to be a measure for how much individuals are disposed to 'adjust their preferences to align them more closely with what they take to be the wishes of others', which is the major requirement for Kitcher's psychological altruism (*EP*, 56).

The main differences between Kitcher's formalization and the two subsequent ones are: a) Kitcher's equation is composed of absolute values instead of functions; b) the left-hand side of Kitcher's equation (v_{Soc}, i.e. actor's social desire) is less explicitly predictive of behaviour than the utility concept in economics (u_1); c) the right-hand side of Kitcher's equation is explicitly meant to describe individuals' *subjective* evaluation of a given outcome (v_{Sol}, v_{Ben}, i.e. actor's solitary desire or desire attributed to the beneficiary) instead of the *objective* value of a given outcome (w_1, w_2, i.e. individuals' welfare or payoff). However, it is easy to transform Kitcher's equation into a functional one without distorting his message.

$$v_{Soc}(\mathrm{x}) = w_{Ego}\ v_{Sol}\ (\mathrm{x}) + w_{Alt}\ v_{Ben}(\mathrm{x})$$

Moreover, if we add the assumptions (usual when building models) that people act upon their desires and evaluate the situation correctly, we end up with the following formula:

$$u_1(\mathrm{x}) = (1\text{-}\beta)w_1(\mathrm{x}) + \beta w_2(\mathrm{x})$$

(where $0 \leq \beta \geq 1$), and the three equations can be directly compared.

From these examples, we can see how difficult it can be to distinguish between objective and subjective preferences, and consequently, between psychological altruism (which represents the actor's real motives for action) and behavioural altruism (which represents the actor's observable behaviour).

3. Examining Kitcher's View of the Evolution of Ethics

Taking a historical perspective, Kitcher claims that ethical practices result from a capacity that has been linearly and gradually acquired on a specific branch of the phylogenetic tree: the one with humans as its most recent shoot. In the early stages, our hominid ancestors displayed altruistic abilities. These basic (mostly emotional) forms of psychological altruism were able to evolve in the first place thanks to the mechanism of parental care. Some great apes, whose ancestors diverged from hominids five million (or more) years ago, also seem to

possess these basic altruistic abilities. Thus some of our present day evolutionary cousins share the essential components of ethics with us, while we humans have developed new features, among which the most important is the capacity for normative guidance. The latter is an essential element for initiation of the 'ethical project': it is the capacity to recognize a rule as a 'command'.

It is worth emphasizing that Kitcher does not conflate the phenomenon of altruism with ethics (*EP*, 134). As he writes, "[a] central theme of my approach to altruism is that there are preethical forms of altruism and that these are realized in animals who have not yet acquired ethical practice" (*EP*, 44). Psychological altruism becomes 'ethically charged' only with the aid of normative guidance. Normative guidance enables altruism to transcend its original domain, originally restricted to close kin, and extend to a broader range.[5] This expansion has an evolutionary rationale for the human species; it is an adaptive response to 'altruism failures' that preclude peaceful living in large communities: "The tension and fragilities of hominid (and chimpanzee) social life arise from the limited altruism of the participants." (*EP*, 222)[6]

Kitcher's scenario is plausible. It is widely accepted among evolutionary theorists that slight changes in genetic makeup can produce important phenotypical effects such as new capacities or behavioural propensities. This is especially true when the slight changes affect neural connections involved in higher order cognitive abilities. Thus, with respect to ethics, humans might be the only species that has undergone the minor but necessary change that has cascaded into the capacity to take an ethical attitude. Possibly, this minor change consisted in recruiting and linking together several preexisting mechanisms as a way of responding efficiently to altruism failures.

However, as Patricia Churchland points out in her latest book (2011, chap. 5), there is a plethora of competing selection theories to explain the origin of human ethical practices, and there is no criteria that allows one to rule any of them out completely. To illustrate this point, consider Kitcher's project of emphasizing the role of psychological altruism. His idea is that "ethical practices grew out of human capacities for psychological altruism" (*EP*, 46): it is the remedy humans have evolved to avoid the failures of a basic form of altruism. As far as I can see, practically, such failures of altruism translate into an incapacity to resolve the everyday difficulties posed by social life in large groups. If the evolutionary challenge posed to humans was to find a solution for facilitating peaceful relationships among growing circles of people, I hardly see the utility of describing this situation *primarily* as a failure of altruism. It might suffice to mention the need to develop more complex forms of mutually beneficial cooperation. Many compatible but possibly independent solutions have been found for realizing this

---

[5] Much in the vein of Peter Singer's 'expanding circle' idea (Singer 1981).

[6] This scenario nicely illustrates the fact that *evolutionary advantage* at the genetic level (expressed by the fact that a gene successfully transmits copies of itself to the following generations) and at the individual level (expressed by the fact that an individual is efficient at producing fully grown offspring) is not to be conflated with greediness or egoism *in the psychological sense*. There is no contradiction in saying that an individual is moved by other-regarding motives (psychological level) *and* that this behaviour is advantageous at some biological level (gene transmission or individual fecundity).

task: the expansion of caring feelings toward wider categories of individuals, the ability to feel some types of social rules as commands, enhanced practical intelligence that makes us aware of the benefits of long-term peaceful interactions, etc. And lately, some philosophers' post-rationalizations have brought these different phenomena together into the broad category of ethical practice.

It is not my purpose to defend this alternative scenario. Rather, this discussion is meant to illustrate that it may be hopeless to attempt to provide a detailed explanation of the evolution of ethics. I tend to agree with Churchland when she writes that "inferring what behavioural traits were selected for in human evolution cannot be solved by a vivid imagination about the ancestral condition plus selected evidence", precisely because this evidence "could be explained in many different ways" (Churchland 2011, 114).

This reservation apart, I am convinced that Kitcher's most valuable contribution to ethics does not rely on the particular historical scenario developed in the first part of his book. The mere recognition of the domain of ethics as a product of evolution is an important advancement in moral philosophy. With such a scientific approach, ethics can be thought of as one (or a series of) ongoing adaptive process endowed with a practical function: to help individuals regulate their social interactions in large groups. It is a fully natural process that has its roots deep in our ancestral history and many of its component-parts might be found in other species (especially, but not exclusively, those that are closely related to us). Thus, descriptive scenarios explaining the origin of ethics in classical philosophical literature (such as the ones proposed by Nietzsche, Rousseau or Hobbes) can be seen in the light of evolutionary theory as the equivalent of ancient geocentric beliefs. The Darwinian turn might not provide answers to all our questions (we have seen that particular evolutionary accounts are based on scarce evidence), but it certainly helps to adopt a constructive scientific view on descriptive ethics which, in turn, impinges on other domains of ethics. Kitcher provides a masterful demonstration of this impact in the second and third parts of his book to which we shall now turn.

4. In Praise of Kitcher's Reasoning Inversion

One of Kitcher's most striking intellectual qualities is his capacity to 'bite the bullet', to take on the consequences of the Darwinian lesson. When needed, he does not shy away from acknowledging implications that contradict well-established and respectable views in moral philosophy. For instance, Kitcher brilliantly captures the devastating implications evolution holds for metaethics. In the light of scientific scrutiny, ethical practice is revealed to be a moving phenomenon intrinsically bound to our evolving biology and culture. This is why the 'ethical project' is a work in progress that constantly needs to be adapted to new emerging biological needs and cultural conditions. This simple knowledge about ethical practices makes it clear that there is no mystery or special quality to ethics, and it smooths the path away from strongly rationalist and realist views in ethics. In Chapter 5, Kitcher refreshingly shakes up conventional philosophical

myths such as the existence of an 'ethical reality' (something we are supposed to have contact with thanks to an ethical faculty), the 'ethical point of view' (a point of view in which people give themselves commands that are not external but somehow their own) or the pure 'moral law within.' No one explains better than he why "appeals to divine will, to a realm of values, to faculties of ethical perception and 'pure practical reason' have to go" (*EP*, 4).

Interestingly, Darwinian metaethical implications also weaken the notion of *ethical truth*, and with it, the way that *ethical progress* is traditionally conceived. Kitcher takes up this topic in Chapter 6 where he provides his most original contribution to ethics. Usually, ethical progress is explained in terms of truth: we achieve progress each time we have gained more insight into ethical truth. This definition does not satisfy Kitcher since there is no ethical truth to 'discover'. He proposes instead to reverse the explanation. "The solution is to revise the concept of progress so it is prior to the notion of truth." (*EP*, 249) Following this line of reasoning, he defines progress in terms of functional efficacy, which is a desirable thing for human beings. Recall that according to Kitcher's evolutionary account, ethical practices have an 'original function': to remedy the failures of altruism that lead to social instability and conflict.[7] Thus, "at its most progressive, the evolution of ethics is (at least as far as its original function goes) a series of responses to the most powerful sources of residual social conflict" (*EP*, 225). On the basis of this new definition of ethical progress, it is then possible to redefine truth: we achieve ethical truth each time we make ethical progress.

There are at least three pleasing upshots of the argumentative inversion suggested by Kitcher. First, it frees ethics from elitism. If ethical progress (and thus ethical truth) is akin to some form of bargaining solution in complex biological and social environments, there is no reason to suppose (which is so often done implicitly under the cover of a respectable rhetoric) that moral philosophers or some particular wise men have privileged access to ethical truth. Second, it facilitates the renewing process in metaethics. Strongly rationalist and realist positions may be given up more readily once it is understood that there is no need to discover truth in order to make ethical progress. Third, and most importantly, it reminds us that the most important task facing ethicists is a practical one: identifying the proper solutions for human failures in everyday social relationships. In the final part of his book, Kitcher makes a crucial step by exploring the normative and practical implications of his theory. He addresses various pressing topics such as scarce food resource management, equal treatment of men and women, and equality of opportunity with regard to novel options offered by biotechnology. One more lesson to take from *The Ethical Project* is that ethicists should not shy away[8] from or dismiss practical questions, an attitude which has produced unfortunate institutional gaps between practical and theoretical ethics in contemporary universities.

---

[7] This desirable aspect of the function of ethical practices (the fact that they contribute to smooth social relationships) confers ethical progress with its normative force.

[8] As a matter of fact, many contemporary ethicists in philosophy departments avoid making practical ethical commitments. This is especially true of evolutionary-minded thinkers who seem intimidated by the tragic history of Social Darwinism (with some notable exceptions, e.g. Rachels 1990; 2004; Gibbard 1990).

Daniel Dennett recently gave a talk in which he proposed three major reasoning inversions that have been helpful in understanding the evolution of human competences.[9] The first reasoning inversion is provided by Darwin when he formulates his theory of evolution by natural selection (Darwin 2003[1859]). Darwin's idea helps explain how complex organisms (including their most intimate mechanisms and faculties) can be produced without knowledge of how to make them—thus without first postulating the existence of an intelligent designer. The second inversion is due to Turing who proposed a similar inversion with respect to computing machines (Turing 1937). With rather simple procedures and techniques, it is possible to build machines endowed with impressive arithmetic abilities, although these machines are completely ignorant of what arithmetic is. Thus, computing efficiency does not necessarily imply an understanding of the task. Putting Darwin and Turing together, it becomes easy to recognize that a large part of human intelligence has been created blindly and produces outputs without any prior understanding of the task to be realized. Our brains, for example, are equipped with a multitude of mechanisms that help individuals to cope with complex environments. These mechanisms highlight or ignore behaviourally relevant information without the mechanisms (and often without the actors themselves) being conscious of this selection process. The third inversion is due to Hume (1993[1772]), who noticed that we commonly interpret an inner anticipation as an external cause of a perception. After having captured associations of events over past experiences, we think that we *see* causation (say we *see* one billiard ball communicating motion to another), when in fact all we see is succession of events (one billiard ball moving before impact followed by a second billiard ball moving after impact). Hume argues that we interpret our subjective expectations formed by associations over time as immediate effects of a perceptible causal link. In other words, we 'project' our impression of causation onto the objects in much the same way we 'project' cuteness onto a baby's face or sweetness onto honey. But there might be no causality, or at least not the sort of causality we think we *see*, between given associated events.

As Dennett writes, these revolutionary reasoning inversions have in common that they "turned everything familiar upside down" (2009, 10061). I propose to add one more thinker to this venerable list: Philip Kitcher, who points to the fact that we wrongly attribute discovery of ethical truth when what we observe is practical ethical progress. Putting Kitcher and Hume together, the result is that the practical ethical commitments we have reached collectively through painful negotiation are felt by us to be necessities. We then wrongly think that we have discovered[10] an external realm of necessary ethical truth or universal moral commands that contain these practical ethical commitments.

Interestingly, one might wonder why Kitcher clings to the notion of ethical truth after having made it so clear that the important aspect of evolving ethics is

---

[9] 'A confusion about consciousness', presented at the opening conference for the new centre of cognitive science at the University of Neuchâtel, 11 January 2012. URL: http://www2.unine.ch/cognition/confcsc (10.02.2012).

[10] In the philosophical literature, there is a variety of ways to interpret this 'discovery': we exert our practical reason, we have the right moral intuitions, we feel the right and the wrong through our moral emotions, etc.

the achievement of functional efficacy (what he calls 'ethical progress'), and *not* the discovery of a truth. Moreover, I have one minor reservation with his analysis (linked to my criticism in the previous section). The function of ethics is not primarily desirable because it remedies the failures of altruism, but because it facilitates social relationships. Since many other cognitive and emotional mechanisms might contribute to realizing this function, I am not sure why we should define ethical progress as the expansion of psychological altruism rather than as the (more general) increased efficiency at regulating one's social interactions in large groups.

5. Concluding Remarks

One interesting lesson to take from Kitcher's book is that we should read Darwin along with Plato and Aristotle in introductory ethics courses. In doing so, we would do well to pay less attention to what Darwin wrote about ethics than to his general theory of evolution. This background scientific knowledge is important for stimulating global reflection on ethics. As Kitcher has taught us, this move would save much intellectual wandering and help us to focus on ethically relevant topics. The most insightful efforts of ethicists do not have much to do with discovering ethical truth or grounding ethical codes that everyone should follow. The truly interesting questions are descriptive and practical ones. First, we need to acquire a better understanding of the practical function of ethics, the conditions of its applicability, and the details of its machinery. Second, on the basis of this factual knowledge, we need to take concrete everyday measures. The very nature of the evolving 'ethical project' demands that we work on it constantly with the aim to put together the right resources and social conditions that best fit our needs as an evolved and highly cognitive species.

Bibliography

Akçay, E./J. Van Cleve/M. W. Feldman/J. Roughgarden (2009), A Theory for the Evolution of Other-Regard Integrating Proximate and Ultimate Perspectives, in: *Proceedings of the National Academy of Sciences* 106, 19061–19066

Binmore, K. G. (2005), *Natural Justice*, New York

Churchland, P. S. (2011), *Braintrust: What Neuroscience Tells Us about Morality*, Princeton

Clavien, C. (2012), Altruistic Emotional Motivation: An Argument in Favour of Psychological Altruism, in: Plaisance, K./T. Reydon (eds.), *Philosophy of Behavioral Biology*, Dordrecht, 275–296

— /M. Chapuisat (2012), Altruism—A Philosophical Analysis, in: *eLS*, 1–7

— / — (forthcoming), Altruism across Disciplines: One Word, Multiple Meanings, in: *Biology and Philosophy*

— /R. A. Klein (2010), Eager for Fairness or for Revenge? Psychological Altruism in Economics, in: *Economics and Philosophy* 26, 267–290

Darwin, C. (2003[1859]), *On the Origin of Species by Means of Natural Selection*, Peterborough

de Quervain D./U. Fischbacher/V. Treyer/M. Schellhammer/U. Schnyder/A. Buck/E. Fehr (2004), The Neural Basis of Altruistic Punishment, in: *Science* 305, 1254–1258

Dennett, D. (2009), Darwin's 'Strange Inversion of Reasoning', in: *Proceedings of the National Academy of Sciences* 106, 10061–10065

Fehr, E./S. Gächter (2002), Altruistic Punishment in Humans, in: *Nature* 415, 137–140

— / — (2005), Human Behaviour: Egalitarian Motive and Altruistic Punishment (Reply), in: *Nature* 433, E1–E2

Frank, S. A. (1998), *Foundations of Social Evolution, Monographs in Behavior and Ecology*, Princeton

Gibbard, A. (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge

Grafen, A. (1985), A Geometric View of Relatedness, in: Dawkins, R./M. Ridley (eds.), *Oxford Surveys in Evolutionary Biology*, Vol. 2, Oxford, 905–907

Hamilton, W. D. (1964), The Genetical Evolution of Social Behaviour, in: *Journal of Theoretical Biology* 7, 1–52

Hume, D. (1993[1772]), *An Enquiry Concerning Human Understanding*, Indianapolis

Kitcher, P. (2011), *The Ethical Project*, Cambridge

Lehmann, L./K. Bargum/M. Reuter (2006), An Evolutionary Analysis of the Relationship between Spite and Altruism, in: *Journal of Evolutionary Biology* 19, 1507–1516

— /L. Keller (2006), The Evolution of Cooperation and Altruism; a General Framework and a Classification of Models, in: *Journal of Evolutionary Biology* 19, 1365–1376

Mayr, U./W. T. Harbaugh/D. Tankersley (2009), Neuroeconomics of Charitable Giving and Philantropy, in: Glimcher, P. W./C. F. Camerer/E. Fehr/R. A. Poldrack (eds.), *Neuroeconomics: Decision Making and the Brain*, Amsterdam, 303–320

Rachels, J. (1990), *Created from Animals: The Moral Implications of Darwinism*, Oxford

— (2004), Drawing Lines, in: Sunstein, C. R./M. Nussbaum (eds.), *Animal Rights: Current Debates and New Directions*, Oxford, 162–174

Singer, P. (1981), *The Expanding Circle: Ethics and Sociobiology*, New York

Stich, S. P./J. M. Doris/E. Roedder (2010), Altruism, in: Doris, J. M. (ed.), *The Moral Psychology Handbook*, Oxford, 147–205

Trivers, R. L. (1971), The Evolution of Reciprocal Altruism, in: *The Quarterly Review of Biology* 46, 35–57

Turing, A. M. (1937), On Computable Numbers, with an Application to the Entscheidungsproblem, in: *Proceedings of the London Mathematical Society* 42, 230–265

West, S. A./A. Griffin/A. Gardner (2007), Social Semantics: Altruism, Cooperation, Mutualism, Strong Reciprocity and Group Selection, in: *Journal of Evolutionary Biology* 20, 415–432