

# Conversations on Human Action and Practical Rationality



Conversations on Human Action  
and Practical Rationality

Edited by

Carlos Mauro, Sofia Miguens and Susana Cadilha

**CAMBRIDGE  
SCHOLARS**

---

P U B L I S H I N G

Conversations on Human Action and Practical Rationality,  
Edited by Carlos Mauro, Sofia Miguens and Susana Cadilha

This book first published 2013

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data  
A catalogue record for this book is available from the British Library

Copyright © 2013 by Carlos Mauro, Sofia Miguens and Susana Cadilha and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-4788-7, ISBN (13): 978-1-4438-4788-9

# TABLE OF CONTENTS

Introduction .....	1
Sofia Miguens and Susana Cadilha	

## **Interviews**

Alfred R. Mele.....	29
Hugh J. McCann .....	51
Michael Bratman .....	85
George Ainslie.....	95
Daniel Hausman .....	119
Joshua Knobe .....	133

Appendix .....	143
References of Project “Conversations on Human Action and Practical Rationality”	



# INTRODUCTION

## SOFIA MIGUENS AND SUSANA CADILHA

The interviews collected in the present volume originated in a research project entitled *Conversations on Human Action and Practical Rationality*, which was conducted at the University of Porto, Portugal, between 2007 and 2011 by MLAG (the Mind, Language and Action Group, a research group of the Institute of Philosophy<sup>1</sup>). The project was designed as a practical project about practical rationality: our aim was to have authors who work on practical rationality and human action answering a set of questions, speaking about their own work and discussing the theoretical differences separating them. The idea was to create an opportunity for some prominent authors in the area to speak for themselves as to what they do assume.

The present volume has the philosophy of action as its domain of reference. Philosophy of action is a field in which issues such as the nature and explanation of actions, the nature of intention, deliberation and decision, the relation between reasons and causation or the possibility of akrasia are discussed. Yet in this book, as was already the case with other activities and publications within the project *Conversations*, what one finds is not so much work on specific issues in the philosophy of action but rather a result of our attempts to explore the direct links between philosophy of action and fields such as moral philosophy, cognitive psychology or the philosophy of economics. Agency and rationality are common denominators to these and several other domains, and that was the true focus of our project<sup>2</sup>.

That Project *Conversations* was conceived as a practical project meant above all that it was designed around interviews, ‘conversations’ on practical rationality and human action, which we intended to pursue with

---

<sup>1</sup> The project was funded by FLAD (Fundação Luso-Americana, <http://www.flad.pt/>).

<sup>2</sup> *Action, Agency and Rationality* is one of the areas of research of MLAG; it has issues of rationality at its core and reaches from philosophy of action to moral philosophy, political philosophy and philosophy of economics.

several people. Thus the script of the interview was particularly important for the whole process. Our script had six introductory questions:

- 1) In your view, what are the most central (or important) problems in the philosophy of action?
- 2) For some or all of the following problems - action, agency and agent - what do they contrast with most significantly?
- 3) Which of these are liable to be rational/irrational?
- 4) In what sense is the thing to do to be decided by what is rational? Are there limits to rationality?
- 5) What explains action, and how? What is the role of deliberation in rationality?
- 6) How is akrasia possible (if you think it is)?

Specific questions about each author's own work were then to be asked, so that the interviews would follow their natural course. These started with:

- 7) How do you think your own work has contributed to the field? What do you consider are your most important contributions? What are your plans for future research?

The idea was that the script would prompt the exploration of each author's thought. Some of the authors whose work we had in mind when we first started considering the script were those interviewed for the present volume: Alfred Mele, Michael Bratman, Joshua Knobe, Daniel Hausman, Hugh McCann and George Ainslie. Among them, some are philosophers, some philosophers with a leaning towards cognitive science, one a psychiatrist and behavioral economist. A longer list of authors helped us delineate the domain of the project: our interest in, and our discussions of, very diverse writings concerning the philosophy of action in contemporary philosophy was our entrance way into the project.<sup>3</sup>

We must reiterate the fact that in the project we intended to address issues as philosophers, but with an eye to conceptual and practical connections with other domains, ranging from cognitive science to

---

<sup>3</sup> In a volume in Portuguese (Sofia Miguens & Susana Cadilha eds., 2012, *Ação e Ética*, Lisboa, Colibri), which was the first publication of the Project, a compilation of more extended references of the project may be found, as well as interviews with several Portuguese philosophers (António Zilhão, Ricardo Santos, João Alberto Pinto, Vasco Correia).



economics, in which questions regarding action, agency and rationality come up. We were interested in locating connections between pure philosophy of action and empirically-minded and experimental work. This was, in fact, the main rationale for our ‘one script’ strategy. Looking at the final results, we believe it was fruitful to ask the exact same questions to philosophers of action (**Alfred Mele** and **Hugh McCann**), to a philosopher of economics (**Daniel Hausman**), to philosophers somehow closer to cognitive science (**Michael Bratman**, who works in the Stanford Symbolic Systems Program; **Joshua Knobe**, appointed in the Program in Cognitive Science at Yale and whose work applies experimental methods to philosophy) and to a psychiatrist who (among many other things) has conducted research on preference reversal in animal behavior (**George Ainslie**).

An introduction to some of the themes of the interviews, at some points referring to passages from the interviews themselves, follows – we built it around a brief profile of each interviewee.

## Delineating the Issues

### Alfred Mele

Alfred Mele has been, since 2000, William H. Lucyle and T. Werkmeister Professor of Philosophy at Florida State University. He works in philosophy of mind, philosophy of action, metaphysics and Greek philosophy; human behavior is the common denominator to his very diverse philosophical interests. He started his philosophical career as an Aristotle scholar, and some of the reasons for his interest in Aristotle, such as his approach to *akrasia*, remained a fixed point throughout his career. For Mele, philosophy of action is a sub-domain of the philosophy of mind with numerous connections to moral philosophy – such connections are visible in topics such as free will, moral responsibility, *akrasia* or motivation for action.

In his first book, *Irrationality – An Essay on Akrasia, Self-Deception and Self-Control* (1987), Mele addressed several problems regarding two forms of irrational behavior: self-deception and weakness of will or *akrasia*. In a later book, *Self-Deception Unmasked* (2001), he would again take on the topic of self-deception, discussing empirical work which he believes supports the positions previously defended in *Irrationality*. In his second book, *Springs of Action – Understanding Intentional Behavior* (1992) he set out to understand what explains actions, focusing on the role of beliefs, desires and intentions in the production of human action. In the

following book, *Autonomous Agents – From Self-Control to Autonomy* (1995), he addressed the very possibility of free or autonomous action, and later, in *Free Will and Luck* (2006), he faced a major theoretical challenge to the idea that sometimes we act freely: the threat posed by luck and manipulation. In his fifth book, *Motivation and Agency* (2003), he developed a theory of the role of motivation in the life of intelligent agents. In his latest book, *Backsliding – Understanding Weakness of Will* (2012), Mele once again deals with akrasia.

Along with akrasia, free will has been a recurrent topic in Mele's work: in another recent book, *Effective Intentions: The Power of Conscious Will* (2009), he examines alleged scientific evidence in favor of the idea according to which free will is an illusion. Such view is supposedly supported by the fact that our brain 'decides' what we do before we are even aware that a decision has been made; there is a small 'time window for free will', of about 100 ms, yet all that can be done in this time window is to veto decisions or intentions. If things are indeed so, our decisions and intentions do not have any role in the production of corresponding actions (e.g., my intention to raise my arm does not have any role in the fact that now my arm is raising) and the idea that we are free agents, and as such worthy of praise or blame, is simply an illusion.

The idea that free will is an illusion is obviously not new in the history of philosophy: suffice it to think of Spinoza or Nietzsche, who, in their works, have done much to deconstruct the phenomenology of the will. Both thought that our sense of making things happen, i.e. our sense of agency, is misleading in that it leads us to think we are masters in our house – 'our house' being our mind and our will. Still, even if that idea is not new, discussing it is, according to people like Mele, made more urgent by the explosion of research on the brain.

But how, exactly, could an attack on free will pose a challenge to the philosophy of action? That could happen in several ways. One may, for instance, doubt that our sense of agency has any actual relation to causation. Notice, as an aid to imagination at this point, that an 'evil genius' scenario does not arise only for the representation of a world outside one's mind – it arises also for the capacity of willing: it might be the case that an 'evil neuroscientist' makes me believe that *I* am doing the things I am doing, when in fact I do nothing for myself since everything I 'do' is controlled by him. Even in the first-person case, we could think it is a mere illusion that there is such a thing as the special role agency is meant to perform in the genesis of what *we* do. *Seeing* something is subject to illusion: for any given case of this, it is conceivable that things should *seem* just as they do where no such thing was seen, or even,

perhaps, where no seeing was going on at all. Are we not also subject to illusion when it comes to (our role in) bringing something about? And why should it not be in the nature of *this* case that such illusion be systematic? This worry aside, to what extent must we *know* what we are doing in order to be doing it? To what extent must one *not* be in the grip of an illusion in order to be acting? And how might such immunity be in the cards?

These are some of the worries about free will arising for a philosopher of action – and Mele, as coordinator of reference books such as *Mental Causation* (with J. Heil, 1993) and *Free Will and Consciousness: How Might They Work?* (with R. Baumeister & K. Vohs, 2010), and currently the leader of a large project on free will, called *Big Questions In Free Will*<sup>4</sup> (2010-2013), has paid much attention to all of them.

In the interview Mele closely examines what is meant by 'action'. One important thing to keep in mind when considering the expression 'human agents' is that *humans do not always act*, i.e. *humans are not permanently and in every circumstance agents*. Here are some of the examples given by Mele in the interview: if an evil neuroscientist uses Sam's brain as a calculator to do multiplication, it is not Sam himself who is calculating when the results of operations occur to him – somehow such results cannot be regarded as his own thoughts. Something occurs in Sam which could be an action of Sam, but in fact is not. If we want to understand what an action is, Mele thinks, it is important to understand why we can and should say that someone like Sam is in such case merely a scenario of events, not a real agent. Or think of Uma: she may calculate something (for instance, a 15% gratuity at the restaurant) using her usual formula of multiplication, or she may do the same calculations compulsively, for no reason – compulsion, in the second case, significantly changes what we want to say about Uma as an agent.

What is at stake here is the fact that something which may seem to be an action – purposeful movements of one's own body, a seemingly intentional sequence of events – is not really an action if the agent 'fails' *to be there*, i.e. if the so-called 'action' merely happens in him, to him, or through him. For something to be an action, it must be *caused* in the right way. For some people, this is a historical matter, regarding what actually causes what: actions must be caused by beliefs, desires and intentions of the agent or they simply will not be actions. That is what the cases above illustrate and such is the outline of a general *causalist thesis*, defended by many authors in the philosophy of action, Mele included. Intentional

---

<sup>4</sup> This is a project sponsored by the John Templeton Foundation.

actions can only be adequately accounted for in terms of beliefs, desires and intentions; and explanations of actions evoking such beliefs, desires and intentions are *causal explanations*.

Once agents (considered as entities which perform certain body movements on the basis of beliefs, desires and intentions) are in place, the question regarding their responsibility for actions arises. According to Mele, only human agents are responsible agents, and, even so, only for some of their actions. This is so because in order that an agent be responsible, it is not sufficient that her action is intentional; self-control is yet another necessary condition (one of Mele's examples here is that of a man who compulsively washes his hands several times a day). Also, it should be considered that the agent might lack the motivation to use self-control. One extra step needed is autonomy; plus, it should be the case that the agent 'could have done otherwise'.

Rationality and irrationality of agents in acting may then be assessed in relation with their *reasons for acting* (e.g., if I throw a heavy vase from the balcony when a man I dislike is passing on the street because I have the desire to kill him and I believe that a heavy vase thrown at his head does the job, I am definitely being rational – yet there is something wrong with me if in the same circumstances I throw a feather or a quail's egg at him). But how is it that an agent ends up having reasons for acting in a certain way? Paths can be diverse, yet a particularly important one is the process of *deliberation*. Deliberating is something like carefully considering what to do. We usually think of deliberations as leading to the formation of intentions, in contrast to a situation where an agent predicts what will be the case with him (if I jump from a building, I will fall – the agent thinks).

Mele sees things here the following way: by deliberating the agent puts forward a decision regarding the practical problem of what to do; this may be seen as an inference process leading to an evaluative conclusion. An agent may deliberate rationally or irrationally; in fact, it is not a necessary condition for the agent to act rationally that the action in question be a direct product of a deliberation. Anyway, if an agent deliberates and ends up thinking that *A is the thing to do*, then she acquires (by default, says Mele) the intention to do A. Practical decisions, i.e., decisions about what to do, are then, according to Mele, momentary mental actions of intention formation; intentions, in turn, are executive attitudes toward plans. Not all intentions, Mele thinks, are actively formed by me in an act of deciding (or in his terminology, not all intentions are 'actionally acquired'): if I open the door of my office every day, when I approach this same locked door today I can 'non-actionally' form the intention to open it.

Such is the way things go by default with agents in what concerns

deliberation and intentional action. Yet there may be conflicts or even clashes between processes of formation of intentions and the agent's motivation – this is what leads to *akrasia*. According to Mele, *akrasia* concerns the relation between an agent's best judgment and her motivation to act, and is basically poor self-control. Typically, decisive better judgments regarding what is to be done are formed on the basis of our evaluation of the objects of our desire. Yet the motivational force of our desires is not always in line with our evaluation of the objects of our desires. It should not be surprising, then, that there are situations where although we think it is better to do A than to do B, we are strongly motivated to do B rather than A. This is what accounts for *akrasia*.

Once one has considered belief-desire mechanisms, deliberation, decision and the motivation of agents, one might think there is still one more question, a global question, as it were, regarding our nature as agents: are we free? Is there such a thing as free-will? As we said above, this is another topic Mele has worked much on, although he doesn't discuss it in the interview. Yet, in a way, that question drives all his work, since he believes that by considering issues regarding desires, motivation, reasons or deliberation the question of free-will is in fact being addressed. Thus, he says: "I myself have no special use for the noun «will»". In this, he contrasts with other authors we interviewed, such as **George Ainslie** or **Michael Bratman**, who do not refrain from saying that in dealing with practical rationality what they are ultimately interested in is a theory of the will. Mele thinks there is no need for such a theory; all that we need in order to render our talk of free-will legitimate is to look closely at all the mechanisms involved.

### Hugh McCann

Hugh McCann is a Professor at the Department of Philosophy of Texas A&M University. His main interest is action – it is action that brings together the topics of philosophy of mind, philosophy of language, metaphysics and philosophy of religion he works on. These topics range from the ontology of events and change to the nature of causality and time, divine action, creation and eternity, and to the problem of evil. He is the author of *The Works of Agency: On Human Action, Will and Freedom* (1998) and *Creation and the Sovereignty of God* (2012). He has also authored numerous articles on quite diverse topics related to the nature of action, e.g. "Divine Sovereignty and the Freedom of the Will", "Volition and Basic Action", "Creation and Conservation", "The Author of Sin?", "Resisting Naturalism: The Case of Free Will", "The Will and the Good"

and “Pointless Suffering: How to Make the Problem of Evil Sufficiently Serious”. He is also the author of the *Stanford Encyclopedia of Philosophy* article on “Divine Providence”. It takes only a quick glance to realize that McCann positions himself against the currently dominant naturalism. In fact, that was the origin of our interest in his work: we wanted to understand how such kind of anti-naturalism would influence a view of action.

McCann stands in a long tradition of connecting problems of action with issues in theology, such as those concerning sin and responsibility, or the relation of human freedom to divine foreknowledge and omniscience. One way to approach his work is precisely, as per his own suggestion in the interview, to try to understand how his way of addressing topics which are central to the philosophy of action (such as, for instance, the nature of basic actions<sup>5</sup>) could bear on the answer to traditional philosophical-theological problems such as the problem of evil.

One very important contrast between McCann and **Alfred Mele** is McCann’s opposition to the causalist approach. McCann’s *volitional theory* is best understood if seen under the light he himself proposes in the interview: tracing the history of the philosophical concept of action from logical behaviorism on, through the causal theory of action, now widespread and widely accepted, he presents his own position, centered on the concept of volition, as a criticism of the latter.

Volition is McCann’s signature-concept as a philosopher of action; it is volition which allows him to explain how action differs from other events and to talk about free will and responsibility. Volition is, according to McCann, the inner activity by which agency is exercised when we perform overt actions. It is not a momentary act, but an activity: the activity of willing. It is volitions that are, according to McCann, basic actions. To fully understand the concept, one might consider an example from the interview (this is in fact a kind of example pervasive in McCann’s writings, since it is crucial for making his case regarding volitions): a patient, suffering from a particular neurological condition, wants to raise his arm, is willing to raise his arm, but to no avail, since his arm does not raise. He is paralyzed, yet he ‘wills’. McCann thinks the theory of action should account for such possibility – there must be some conception of what the case is in such situation. McCann sees nothing wrong with the idea that an action undertaken for a reason is intentional, nor with the idea

---

<sup>5</sup> I.e. those actions which do not take place by means of the agent doing something else (in contrast with non-basic actions such as firing the gun by pulling the trigger, or pulling the trigger by moving one’s finger).

that beliefs and desires of the agent are reasons for action; all he wants is that such ideas do not close our eyes to the fact that *acting is mainly willing to act*, wanting to make something happen. In other words, the essential thing about action is not what happens subsequently but the fact that an agent produces it by willing (to do something). Beliefs and desires may be necessary conditions for an action to take place, but only the volitional act is a sufficient condition. The phenomenological aspect of this is very important for McCann – he thinks that to engage in volition is to feel spontaneously active, not something that merely happens to us.

It is important to understand the way McCann sees the relation between volition and intention, as well as the relation between beliefs, desires and intentions. First, for McCann, intentions cannot simply be combinations of beliefs and desires. Here is another example from the interview: I may want to see Rome, and I believe I can do it by going on a holiday to Italy this summer, yet it hardly follows from this that I intend to go to Italy this summer. What is missing? What would be necessary for me to intend to go? What is missing is intention – intending to do something requires that I have decided, that I am committed to doing something and that I actually form a plan for it.

But if we have intentions only, what we have are states, not events, and actions are supposed to be events. On the other hand, if we think, for instance, about a movement of one of our limbs (for instance, our arm's upward motion), what we have is an event but not, by itself, an action (it may either be an action or not). This is the space filled by *volition*, the means by which we execute intentions. An intention, being a state, says McCann, cannot execute itself; that is why volitions are needed. Another important difference here, then, is the difference between desire and volition. A desire is always something which befalls us, something we find ourselves having, while a volition is, in McCann's term, 'actional', it is itself active, not the sort of thing that ever befalls a person.

Even if moral and theological implications of the discussion of action are, obviously, of great interest to McCann – he is interested in issues such as the duties and obligations of God and men, in why God created a world in which there is suffering and sin, as well as in questions of guilt – he is careful to emphasize in the interview that his views leave open the question whether the will is caused, either by reasons or by intentions. Are we free agents? If the will is subject to causation can we possibly be free? For McCann the problem of free-will is one of the problems of philosophy of action, a problem he sees as concerning the forming of intentions (not executing them). He is an incompatibilist, a libertarian who believes having free-will *is* forming intentions. The libertarian must argue that

reasons have explanatory power without causally determining the formation of intentions; he also has to explain why a decision is not an accident for which the agent is not responsible. A compatibilist, in turn, must find a way to distinguish between free and non-free actions in a deterministic world, and to offer a theory of responsibility and guilt that is compatible with determinism. Neither has a simple task. This seeming impasse between libertarians and compatibilists leads McCann, who is himself a libertarian but who considers the notion of agent causation a vacuous notion, to evoke a practical perspective on action, that of the phenomenology of acting. As he puts it in the interview, “From the practical perspective of the agent (...) it is conceptually impossible for decision or willing to count as ‘irruptions’, because it is conceptually impossible for these phenomena to occur accidentally or inadvertently. To see this, simply imagine the example I mentioned before: how you would react if a student came up to you and said, «I’m sorry to have missed your lecture yesterday Professor (fill in your name). I accidentally decided not to come»”.

Finally, in what respects *akrasia*, McCann thinks that it happens frequently and that it is not a particularly problematic issue for the philosophy of action – *akrasia* simply happens when agents lack the motivation to carry out the intentions they formed. Contrary to Davidson’s famous proposal according to which there is nothing specifically moral in the problem of *akrasia*<sup>6</sup>, McCann does not flinch at the idea that the concept does have moral connotations – in his opinion *akrasia* is often a failure of the righteousness (*fortitude*) of a person and as such says something about her character.

### Michael Bratman

Michael Bratman is Durfee Professor in the Department of Philosophy at the School of Humanities and Sciences at Stanford University. Stanford University is an important place for the history of cognitive science in the second half of the twentieth century, and Bratman himself is close to its *Symbolic Systems Program*, a program of cognitive science which includes several disciplines, from philosophy to artificial intelligence, whose goal is the study of computers and minds or, in other words, the study of natural and artificial systems which use symbols to represent information. His philosophical work led him to develop a model of human practical reasoning, the BDI (Belief-Desire-Intention) model, which is used in many

---

<sup>6</sup> Cf. D. Davidson, “How is Weakness of the Will Possible?” In *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.



areas, including artificial intelligence. In 2008 he received the IFAAMAS (International Foundation for Autonomous Agents and Multi-Agents Systems) Award for Influential Papers, an award marking key contributions to research on agents and multi-agent systems.

How does a philosopher end up inspiring work in artificial intelligence in such a direct way? Bratman believes the Western philosophical tradition has focused mainly on mind and knowledge rather than on agency. Yet agency, in the sense of ‘behavior potentially under the control of an oriented mechanism, which pursues a particular purpose’, is a concept common to a large number of disciplines, for all of which philosophical studies of agency are of interest. Such disciplines range from artificial intelligence, and cognitive science in general, to the social and political sciences. Bratman’s first book, *Intention, Plans, and Practical Reason* (1987), with its focus on *intentions* and *plans*, goes a long way in answering the question of how a philosopher ends up inspiring work done in artificial intelligence and having a software model developed to program intelligent agents inspired by his investigations. The book had a strong impact, calling attention to the complex and constitutive role of *intentions* and *planning* in rational agency. Intending, as it interests Bratman, is not just intention in action, the kind of phenomenon people such as Elizabeth Anscombe and Donald Davidson paid most attention to – it is also planning, i.e. developing stable and future-directed intentions. Moreover, according to Bratman, the phenomenon of *intending to act* is not just the result of practical reasoning: it is also a fundamental element of the information input of agents. Understanding intentions and plans is essential to understand the nature of agency and a key to characterize the crucial difference between practical rationality, concerning what to do, and theoretical rationality, concerning what to believe.

The BDI model is in fact a proposal about the nature of practical rationality, according to which at its centre lies the phenomenon of intending to act; it separates two distinct capacities, the capacity of selecting a plan and the capacity of implementing current plans; according to it, agents are able to balance the time they spend deliberating on plans and implementing plans. Bratman continued his investigation in *Faces of Intention – Selected Essays on Intention and Agency* (1999), and *Structures of Agency: Essays* (2007), exploring how characteristically human agency is related to issues of self-determination, self-government and autonomy. In fact, according to Bratman, the issues at the core of philosophy of action concern the basic structures of *human* agency, and his work has extended from questions regarding the nature of plans and intention to specific phenomena such as *shared intention* and *shared agency*. These are, he believes, particularly important for understanding

human coordination and sociality.

His work on *shared agency*, *shared intention* and *shared valuing* has been influential in the philosophy of law and political theory, and naturally so, since it aims at characterizing basic human capabilities, such as those of sociality and self-government, without reference to which there could be no such fields. These capabilities are indeed decisive in making us human, and according to Bratman the most crucial of them is shared intention.

Bratman regards his *planning theory* as a 'modest theory of the will' - in fact, he sees it as an alternative to other theories of the will, for example **George Ainslie's**. According to Bratman, and contrary to what incompatibilists take it to be the case, the will is not something extraneous to the causal order: creatures with a 'will' are simply creatures endowed with certain psychological structures, namely capacities for imposing a structure on their thought and action. The question is to understand such structures, and Bratman thinks understanding the nature of *planning* and associated norms is crucial there. Given the fact that human lives are extended in time and socially coordinated, and given the fact that this happens in a context of cognitive limitations and limited access to information, *planning* is fundamental (it's worth keeping in mind here, in order to better place Bratman in the history of cognitive science, that he was deeply influenced by Herbert Simon's theory of *bounded rationality*<sup>7</sup>).

Acts of choosing and deciding are the standard ways we use to delineate plans in our minds. Human agents are *planning agents* and Bratman insists that planning agents are a very special type of *purposive agents* (this last category encompasses many other animals, e.g. rats or bats or cows). Such purposiveness, which characterizes all agents as agents, is not to be identified with the phenomenon of planning that Bratman is interested in. In planning theory, *intentions* have a special relevance, in contrast with classical philosophical discussions of human motivation to act, where noticeably more attention is paid to desires. Even in 20<sup>th</sup> century philosophy of action comparatively little attention was paid to intention in Bratman's sense. In the interview, Bratman says he sees his work, and its focus on planning, as resulting from a discussion about agency and intention between Anscombe and Davidson which has never actually taken place. He directs our attention to the many points of

---

<sup>7</sup> Economist and cognitive scientist Herbert Simon (1916-2001) put forward the idea of bounded rationality in contrast with conceptions of rationality as optimization (one example being the rational choice theory's idea of maximization of expected utility). The idea of bounded rationality intends to capture the doings of a rational agent in situations of limited access to information which are quite common in the real world.

agreement between them: both admit 'multiple descriptions' of what happens and they agree on the topic of individuation of actions. The classic points of conflict between the two concern causal explanations of action (Davidson defends them, Anscombe does not), and the non-observational knowledge which agents such as ourselves have of themselves while acting (something upon which Anscombe insists much and which is absent in Davidson). Yet, according to Bratman, there is a kind of tacit agreement between Anscombe and Davidson about *intending to act*: none of them thought that going that way would be very fruitful. But this is precisely the way to go, according to Bratman. Even if both intentions and desires are, in Davidson's terminology, 'pro-attitudes' (mental attitudes related to action), intentions are special in that they involve *commitment*. It is such commitment which makes persistence of plans possible, and which makes new plans possible based on those with which the agent is already committed. Also, intentions are, in contrast with desires, subject to requirements of rationality; together with beliefs, intentions have the function of providing a consistent model of the future of the agent, for the agent. As Bratman puts it in "Intention, Belief, Practical, Theoretical": "Central to the planning theory is the idea that intentions – in contrast with ordinary desires – are both embedded in characteristic regularities and are subject to distinctive rational pressures for consistency and coherence. There is, in particular, a rational demand that one's intentions, taken together with one's beliefs, fit together into a consistent model of one's future. There is, further, a rational demand that one's intentions be means-end coherent in the sense, roughly, that it not be true that one intends E, believes that E requires that one intend means M, and yet not intend M. And these norms of consistency and coherence are operative in a planning agent's practical reasoning." Understanding the nature of those *rational requirements* is a central concern of Bratman. And to characterize intentions this way is to see them as elements of future-directed plans, which are central to practical rationality.

Yet another reason for the great interest Bratman's work has for the studies of rationality and human action is his exploration of the epistemology of practical rationality, which, as he shows, contrasts with theoretical rationality in many interesting ways. Bratman focuses for instance on the fact that plans are typically partial, i.e. they have to be filled in by reality and context, and also on the fact that in contexts of practical decision-making a cognitive attitude of 'acceptance in a context' takes over (which contrasts much with belief<sup>8</sup>).

---

<sup>8</sup> Cf. M. Bratman 1992 'Practical Reasoning and Acceptance in a Context', *Mind* 101, 1-14.

## George Ainslie

George Ainslie is an American psychiatrist, psychologist and experimental economist and the author of two books which were especially important in *Project Conversations: Picoeconomics – The Strategic Interaction of Successive Motivational States Within the Person* (1992) and *Breakdowns of the Will* (2001). The idea of 'picoeconomics' is the following: the same way classical economics aims at describing negotiation for limited resources between institutions and microeconomics aims at describing negotiation for limited resources between individuals, picoeconomics aims at describing *interactions for the control of the finite behavioral capacity of an individual within that individual*. As in the other two cases, the *interactions for the control of the finite behavioral capacity of an individual within that individual* resemble a negotiation between parties. In *Breakdowns of the Will*, Ainslie applies the picoeconomics approach to a model of the self (or person) and of the will. This is what allows him to explain the so-called 'collapses' of the will.

Ainslie's medical background clearly marks his approach: he always has in mind not only cases which are the classic object of interest of philosophers, such as more or less abstractly conceived cases of akrasia, but clinical conditions, namely addiction (to drugs or alcohol, for example, but many other types of addiction as well). His proposals regarding the nature of action, self and will, are in fact intimately related to his studies of addictive behavior, with which he dealt very closely in his clinical work<sup>9</sup>. He actually believes that the ambiguity which characterizes addictive behavior (the willing and not willing to do what one sets about to do, as in the case of the addict who wants and does not want the heroine he is about to take) is quite pervasive even outside clinical contexts.

What is particularly relevant for understanding the self and the will in the phenomenon of ambiguity is the fact that agents are very often seduced by urgent pleasures with destructive objects, thus interfering with their own long-term stable choices. It is as if there was a self always alert to the possibility of immediate, strongly motivating, compulsive, pleasure – such as in 'I want these drugs now' – and such self often won and controlled action, even if that meant failing to give satisfaction to the long term interests of another self that is also oneself, and failing to respond to what that other self is committed to. If Ainslie is right in thinking that this is a particularly important structure for the study of action and agents, that means that understanding the preferences of agents such as ourselves

---

<sup>9</sup> He was Chief Psychiatrist in the Veteran Administration Medical Center, Coatesville, Pennsylvania.

involves much more than the computations of relative values considered by rational choice theory. A basic idea of Ainslie is that different parts or aspects of personality are in conflict in individuals (this is something he explicitly relates to the Freudian theory of the *id*, the *ego* and the *superego*) and that has to be considered always, in choice behavior. This is particularly clear in addictive behavior, thus making addictive behavior particularly significant in the study of action.

Ainslie's ideas spread to behavioral economics, especially through Richard Thaler's<sup>10</sup> theory of multiple selves. Also, together with Drazen Prelec's<sup>11</sup> investigations, Ainslie's work was one way through which studies on operant conditioning joined other approaches to decision making, creating a challenge to mainstream economic thinking, centered on rational-choice theory.

In the interview, again evidencing his medical and clinical background, Ainslie puts forward a definition of action quite different from those of the other authors' interviewed: for him actions are *teleological processes, which are repeatable and reward-responsive*. Among the examples he gives are bulimia and other kinds of addictive behavior. It is the 'reward-responsiveness' of actions he focuses on and it is within such framework that he answers our questions regarding choice, deliberation, akrasia, etc.

Since he does not recognize a distinction in kind between actions and passions, or between actions and thoughts, but only degrees of motivatedness and degrees of deliberateness, Ainslie cannot simply classify actions into rational and irrational. He believes that the test for rationality which comes closest to the ordinary meaning concerns whether a choice serves the long range interest of the self. In approaching the issue, he prefers to focus on motivation, which, he says, implies a universal currency, best called 'reward', which is necessary to settle the competition between members of the class of processes that can be substituted for one another. He refers to the hypothetical space at which these processes compete for expression as the *motivational marketplace*.

Speaking about 'reward' in this way is a mark of Ainslie's experimental investigation of operant conditioning in animals. Ainslie has worked with psychologist Howard Rachlin<sup>12</sup>, and investigated, in particular, the phenomenon of inter-temporal choice in pigeons. He was first to demonstrate experimentally the phenomenon of preference reversal in agents who have immediate benefits in view, a phenomenon he explained

---

<sup>10</sup> American economist, b. 1945, University of Chicago Booth School of Business.

<sup>11</sup> Professor of economics in the MIT Sloan School of Management.

<sup>12</sup> American experimental psychologist, Emeritus Professor at SUNY.

in terms of *hyperbolic discounting of prospective rewards*. This means that the agents' valuing of prospective reward stands in inverse proportion to its delay (this is an idea stemming from Rachlin's work). Ainslie then integrated such views into experimental and theoretical work on inter-temporal choice. In fact, the term 'picoeconomics' is often used to describe the implications of one specific experimental discovery: the tendency people have of showing strong preference for immediate payoffs in alternative to long-term benefits, a tendency which is stronger the closer both situations are to the present moment. In other words, given two (comparatively) similar rewards, humans show a preference for the one which arrives earlier. For example, many people, when offered a choice between getting \$50 now and \$100 a year from now choose \$50 now. However, given a choice between \$50 in 5 years, and \$100 in 6 years, they all choose \$100 in 6 years, even if this is the same choice only seen at a greater distance. It is thus said that humans discount the value of the later reward by a factor which increases with the increasing delay of the benefits. A large number of experiments confirmed that spontaneous preferences of human and non-human subjects followed a hyperbolic curve rather than the conventional exponential curve which would reflect consistent choice over time.

Ainslie's book *Picoeconomics* covers such topics, but one can also find there many insights derived from the philosophical tradition of reflection on the passions, a tradition which connects the Stoics, David Hume and (very centrally) Freud. Anyway, a central point of Ainslie's is that a formal way of thinking about choice, such as the utilitarian way of thinking about choices in terms of maximization of expected utility, taking as a touchstone simple monetary choices, simply does not capture the complexity, nor the ambivalence, of most human choices.

It is in his book *The Breakdowns of Will* that Ainslie deals more directly with akrasia. The idea of *hyperbolic discounting of prospective rewards* suggests a model of the self as a population of processes in search of reward – the short-term and long-term interests of agents are seen as reward-seeking processes competing with each other, according to a rule of maximization of expected rewards discounted at every moment of choice. This is combined with an inter-temporal bargaining model of the will. In the context of such competition, the 'will' is not (contrary to what rational-choice theory holds) a superfluous concept: it has a crucially important function: maintaining the agent's preference for LL [larger and later], as opposed to SS [smaller and sooner].

As for human action *par excellence* (to use D. Velleman's term), how is it that, according to Ainslie, consciousness, deliberation or freedom

come to be? Ainslie believes the question should be posed in terms of the relation between motivation and deliberation: “Most of the processes that seem to have been discussed by philosophers are not only motivated but also deliberate”, and he thinks that deliberate actions have been tested in the marketplace of reward; in fact one chooses them while conscious both of them and of their immediate alternatives. In this sense a deliberate action is one chosen “all things considered”.

Still how could such collection of processes be a person, and ‘one and the same person over time’, as Locke would put it? According to Ainslie a (single) person comes into existence out of populations of competing interests: “These properties of action permit a theory of how a person, that is, a population of reward-seeking processes, can form higher mental processes – ‘ego-processes’ and an ‘ego-identity’.” A person is thus a single entity that extends over time and who knows that she cannot always count on herself: temptations come by, and it is very possible that she succumbs to them in the future. What she has to do then is ‘negotiate with herself’ such inter-temporal relations and preferences. In the interview Ainslie evokes Jon Elster’s emblematic case of Ulysses and the Sirens<sup>13</sup>: Ulysses does not want to be tempted by the Sirens, yet since he believes he will in fact be tempted by the Sirens, he decides to bind himself to the mast, planning his action in order to keep himself from succumbing to temptation when the time comes.

With his theory of the self and the will, inspired by behaviorist psychology on the one hand and by Freudian themes on the other, Ainslie makes clear not only the difference the biological nature of (at least some) agents might make, but also that a model of the self is crucial when addressing issues of practical rationality.

### **Daniel Hausman**

Daniel Hausman is Herbert A. Simon Professor in the Department of Philosophy of the University of Wisconsin – Madison, and a prominent name in the philosophy of economics. For people with a specific interest in the philosophy of science, economics is a particularly exciting subject: it shares many features with the natural sciences, including the use of mathematical methods, while its objects are social phenomena. Also, economic theories such as theories of welfare and social choice often involve substantive (moral, political) philosophical commitments. Work in

---

<sup>13</sup> Jon Elster, *Ulysses and the Sirens*, Cambridge, Cambridge University Press, 1979.

the philosophy of economics is thus bound to require quite diverse theoretical skills and interests, and Daniel Hausman's education path is a good example of that. He attended Harvard College majoring first in biochemistry and then in History and English Literature. After teaching intermediate school in the Bronx and obtaining a Master of Arts in Teaching at NYU (New York University), he spent two years studying moral science at Gonville and Caius College at Cambridge University, in the United Kingdom, completing a PhD in philosophy at Columbia University, New York, in 1978.

Hausman's research has focused on methodological issues, as well as on metaphysical and ethical issues arising at the borders of economics and philosophy. In collaboration with Michael McPherson he founded the journal *Economics and Philosophy*. He was editor of the journal for the first ten years of its existence. He has also coordinated an important anthology, *The Philosophy of Economics* (2007), and published several books, namely *Capital, Profits, and Prices: An Essay in the Philosophy of Economics* (1981), *The Inexact and Separate Science of Economics* (1992), *Economic Analysis and Moral Philosophy* (co-authored with Michael McPherson, 1996), *Causal Asymmetries* (1998), *Economic Analysis, Moral Philosophy and Public Policy* (co-authored with Michael McPherson, 2006) and, more recently, *Preference, Value, Choice and Welfare* (2011).

In his *Stanford Encyclopedia of Philosophy* article on "Philosophy of Economics", Hausman maps the problems of philosophy of economics, characterizing the philosophy of economics in an illuminating way as consisting of three main types of investigations: (a) investigations of rational choice, (b) investigations on the appraisal of economic outcomes, institutions and processes, and (c) investigations on the ontology of economic phenomena and the ability to acquire knowledge of them. According to Hausman, although these inquiries intersect in many ways, it is useful to keep such branching in mind, especially since the branches can be seen respectively as (1) a branch of action theory, (2) a branch of ethics (or normative, social and political, philosophy) and (3) a branch of philosophy of science.

In the interview, Hausman begins by bringing to our attention that the very identification of the main problems of philosophy of action depends on the interest with which one approaches the field. In his case, having economics as a background and rational choice theory as a reference, the major problems are the nature of preferences, the difference between preferences and value judgments and the connection between preferences and wants, beliefs, desires, choices and action (or, in other words, the relation between rational choice theory concepts, as used by economists,



and common descriptions of human action, closer to common sense psychology).

It is hardly disputed that rational choice theory involves a specific perspective on practical rationality. For example, when asked about the role of deliberation in rationality, Hausman pointed out that this is itself controversial: in economics it is assumed that agents have completed their deliberations, that they already have a definite and clearly fixed ranking of preferences. They then choose according to this ranking – that is what acting rationally is. In such conditions, to deliberate is simply to calculate – nothing more needs to be said about deliberation. Eventually, at a later stage, problems concerning the formation or the change of preferences may arise (or even questions about the very relevance of the notion of ‘preferences’ – Oxford philosopher John Broome, for example, would prefer to simply speak of ‘goods’, so as not to skew or bias questions in advance<sup>14</sup>). But such reflections mostly do not come up in economists’ work.

A bit more should be said about agents and the ranking of preferences, given the fact that, from the perspective of economics, that is what rational agents are: entities characterized by a ranking of preferences, acting rationally when choosing according to it. In the *Stanford Encyclopedia* article, Hausman, while analyzing the history of economic thought from Aristotle, through the Physiocrats, David Hume and Adam Smith, to the present, calls attention to the fact that 20<sup>th</sup> century economists have generally abandoned earlier hedonistic formulations of choice and preference, which were closely linked to utilitarian philosophy and which focused on the happiness of agents. It was in the place of such formulations that talk of a ranking of preferences became common. Rational agents are characterized by their preferences; preferences are rankings of objects of choice. To deliberate is to create a ranking of the alternatives we face. It then becomes a crucial feature of rational agents that they are able to make a *consistent* ranking. This means that rankings are *complete* – for two alternatives *x* and *y*, the agent either prefers *x* to *y*, or prefers *y* to *x*, or is indifferent. Also, the rational agent's preferences are *transitive*. Thus we have the economist's picture of the rational agent: he or she has complete and transitive preferences and armed with such preferences he or she chooses among alternatives. This angle of approach is in fact essential to distinguish economics from other social sciences: what is distinctive of economic investigations of social reality is supposed to be precisely the (rational) nature of the preferences of economic agents.

---

<sup>14</sup> Cf. John Broome, *Weighing Goods*, Oxford, Blackwell, 1995.

Admittedly, something like desire for wealth and consumption should be present behind such rational preferences – this, however, is simply assumed. Basically, wealth and consumption are what makes humans happy; preferences such as the ones of the ascetic man are, as Hausman comments in the interview, quite 'strange'.

What we have here is the outline of a theory of rationality, and thinking about the nature of such theory in its multiple branches and domains is part of the occupation of philosophers of economy. Still one may argue, for example, that this is too weak a theory, as it says nothing about beliefs, or about what rationality implies when agents do not know everything relevant to the choices they make. One might also consider it too strong: one might for instance argue that there is nothing irrational in having incomplete preferences in situations which involve uncertainty – situations which definitely abound in the lives of agents such as us. And it may be reasonable to suspend judgment or not to choose among alternatives whose nature is not completely understood – why should such a stance be considered irrational?

Moreover, as for the transitivity of preferences, while it does seem a plausible requirement of rationality, it also seems to face abundant experimental evidence that people's preferences are not in fact transitive. Arguably this doesn't imply that agents themselves are irrational, since it is only a part of their behavior that could be characterized in that manner. Anyway, the interpretation of data regarding change of preferences and 'irrational' choices is a core issue in studies of rationality, calling for the combination of abstract models with empirical research.

Extending the theory of rationality to circumstances involving risk and uncertainty calls for yet more principles and technical instruments, such as Bayesian ones, whose nature is also up to the philosopher of economics to analyze. In any case, it is important to keep in mind that rational choice theory involves the *formalization* of the conditions of rationality, allowing for formal manipulations. If we have agents with complete and transitive preferences who also satisfy an additional condition of continuity, then they can be represented by the ordinal utility function. One can define a function that represents the preferences of the agent such that  $U(X) > U(Y)$  if and only if the agent prefers  $X$  to  $Y$  and  $U(X) = U(Y)$  if and only if the agent is indifferent between  $X$  and  $Y$ . The function represents the ranking of preferences only – it contains no additional information. Any transformation of "U" that preserves the order also represents the preferences of the agent. When, in addition to that, the agent's preferences satisfy the condition of independence, and other technical conditions, they can be represented by the expected utility function. This is a function

which has important properties much discussed by economists and philosophers of economics.

The above is just an elementary sketch of what goes on in the philosophy of economics. Connections between philosophical issues about action, agency and rationality and the field of economics were at the very origin of project *Conversations*<sup>15</sup>, and Daniel Hausman's work helped us making such connections clear.

### Joshua Knobe

Joshua Knobe is currently a Professor in the Cognitive Science Program at Yale University and a major figure in experimental philosophy. His case for experimental philosophy has had plenty of coverage in mainstream and online media, from *The New York Times*, to *Slate* or *bloggingheads.tv*. Experimental philosophy advocates setting aside philosophers' appeal to (their own) intuitions trying instead to understand how ordinary people think by means of empirical studies.

Knobe's name is associated with the "Knobe Effect" or "Side-Effect Effect". The "Knobe Effect" emerged in a much-discussed study of intentional action<sup>16</sup>. In the study Knobe confronted people on the street (people spending time in a Manhattan public park) with the following scenario (each subject read a vignette): "The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but will also harm the environment.' The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed." The subjects were then asked how much blame the chairman deserved for what he did and whether he had *intentionally* harmed the environment. A vast majority of people (82%) said the chairman harmed the environment intentionally. This is the 'harm condition'; in the 'help condition', the vignette is exactly the same, except the word 'harm' is replaced by 'help'. Thus, subjects read: "The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but will also help the environment'. The chairman of the board answered, 'I

---

<sup>15</sup> Carlos Mauro, one of the editors of this book, and an economist by training, who did a PhD in Philosophy on the topic of rationality in action (Porto, 2009), was instrumental in conceiving of Project *Conversations on Practical Rationality*.

<sup>16</sup> Cf. Joshua Knobe, 2003, "Intentional Action and Side Effects in Ordinary Language", *Analysis* 63, 190-193.

don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped." Asked whether the chairman had intentionally helped the environment, a majority of subjects (77%) said 'no'.

The asymmetry between the 'harm' and the 'help' scenarios is known as the "Knobe Effect". Based on it, Knobe has argued that so-called folk-psychological attributions of intentionality are not morally neutral, and that, if such is the case, one should not regard them as neutral tools for predicting and explaining behavior as is often done in philosophy and cognitive science. Instead, the common concept of 'intentional action' should be regarded as something like a '*multipurpose tool*', with different cognitive uses.

The example above is not in itself a particularly controversial piece of experimental philosophy; more needs to be said about the reasons why experimental philosophy became so controversial within academic philosophy<sup>17</sup>. Even if much of what is actually done in experimental philosophy can be regarded as cognitive science research on topics ranging from intention and consciousness to free-will and the emotions, the fact is that experimental philosophy began with an extra purpose: to militate against traditional armchair philosophy. Experimental philosophers claim that in order to understand the ways humans think and act, philosophers should 'go out and run empirical tests', rather than stick to their own intuitions. This is the reason why experimental philosophy triggered a heated debate about what one does when one does philosophy.

Now, 'traditional' philosophers who are less than enthusiastic about experimental philosophy mostly think there is nothing wrong with the empirical research of the topics referred to above – what they dispute are the reasons why anyone should consider that Knobe, or any other experimental philosopher, is doing philosophy, as opposed to social or cognitive psychology. The controversy around experimental philosophy can easily be reconstituted by reading the Manifesto Knobe co-wrote with Shaun Nichols<sup>18</sup>, and critical responses to it. Some such responses are Antti Kauppinen's ("The Rise and Fall of Experimental Philosophy"<sup>19</sup>) or

---

<sup>17</sup> The following paragraphs express the views of the authors of the present Introduction, and not those of Carlos Mauro, who is himself a practitioner of experimental philosophy.

<sup>18</sup> Cf. Joshua Knobe & Shaun Nichols, *Experimental Philosophy*, New York, Oxford University Press, 2008.

<sup>19</sup> Antti Kauppinen, 2007, "The Rise and Fall of Experimental Philosophy", *Philosophical Explorations* 10 (2): 95-118.

Timothy Williamson's, in his book *The Philosophy of Philosophy*<sup>20</sup>. Many other comments can be found on the Web. Antti Kauppinen ends his article by saying that "At best, survey results provide food for thought – but we are better nourished if instead of designing artificial setups we pay close attention to what is said in real-life situations of language use, as conscientious philosophers have done at least since Socrates". Among other things, this is a methodological observation about the actual role of language in philosophical investigations. In fact, something that might exasperate armchair philosophers (i.e. those analytic philosophers whom experimental philosophers see as representatives of the 'technical mainstream') is the fact that experimental philosophers (at least in the Knobe-Nichols Manifesto) characterize the 'method of conceptual analysis' as consisting in attempts to 'identify precisely the meaning of a concept by breaking the concept into its essential components'<sup>21</sup>. This is a surprisingly simplistic description, which basically overlooks the whole history of analytic philosophy since its late nineteenth century beginning with Frege. Not only it skips the role logic has had in investigating the nature of thought, as a response to scientism and psychologism rising in late nineteenth century, but also disregards the fact that philosophical analysis of thought simply is not, for the founders of the analytic tradition, analysis of concepts in people's minds. Yet the fact that at the origins of analytic philosophy logic and language did come together in a conception of philosophical method (the philosophical significance of this is obviously open to discussion<sup>22</sup>) seems to go unnoticed by experimental philosophers: for them 'method' seems to denote merely *experimental* method, the only method which can be 'scientific'.

Still, an experimental philosopher could argue that the dispute is not only, or mostly about method but about issues: in the interview, Joshua Knobe stressed the importance not only of the methods of experimental philosophy for the philosophy of action, but also the fact that experimental philosophy is reviving classic topics of philosophy, such as those addressed by a philosopher like Nietzsche, topics concerning how human beings really are, in contrast with the abstract technical issues of armchair

---

<sup>20</sup> Timothy Williamson, *The Philosophy of Philosophy*, Oxford, Blackwell, 2007.

<sup>21</sup> Cf. Knobe & Nichols, 2008, p. 4.

<sup>22</sup> In fact it is historically more accurate to see the history of analytic philosophy since its fregean beginnings as an ongoing dispute around several conceptions of method, all having logic and language as their reference. Such dispute has to do, namely, with diverging conceptions of the relation between formal languages and natural languages, which obviously bears on the conception of what an 'analysis' by means of logic does.

philosophers. One example Knobe gives of such Nietzschean questions is that of why people believe in free-will (the answer might lie, according to Nietzsche, in our desire to justify acts of punishment: Nietzsche speaks of the 'metaphysics of the hangman').

As stated in the Manifesto: "The ultimate hope is that we can use this information to help determine whether the psychological sources of the beliefs undercut the warrant for the beliefs. The basic approach here should be familiar from the history of philosophy. Just take a look at nineteenth-century philosophy of religion. At the time, there was a raging debate about whether people's religious beliefs were warranted, and a number of philosophers (Marx, Nietzsche, Feuerbach, etc.) contributed to this debate by offering specific hypotheses about the psychological sources of religious faith. These hypotheses led to an explosion of further discussion that proved enormously valuable for a broad variety of philosophical issues. But then something strange happened. Although arguments of this basic type had traditionally been regarded as extremely important, they came to occupy a far less significant role in the distinctive form of philosophy that rose to prominence in the twentieth century. The rise of analytic philosophy led to a diminished interest in questions about, for example, the fundamental sources of religious faith and a heightened interest in more technical questions that could be addressed from the armchair. The shift here is a somewhat peculiar one. It is not that anyone actually offered arguments against the idea that it was worthwhile to understand the underlying sources of our beliefs; rather, this traditional form of inquiry seems simply to have fallen out of fashion. We regard this as a highly regrettable development. It seems to us that questions about the sources of our religious, moral, and metaphysical beliefs are deeply important questions and that there was never any good reason to stop pursuing them. Our aim now is to return to these questions, this time armed with the methods of contemporary cognitive science."<sup>23</sup>

One thing that could be said about this is that experimental philosophers seem to be aware of Anglophone philosophy only. If they were aware of a wider philosophical tradition it would be obvious to them that such topics, as well as Nietzsche himself as a major philosopher, were in fact never abandoned. Yet what continental philosophers inspired by Nietzsche (such as Foucault or Deleuze) aim at is certainly not in the least similar to the experimental philosophers' proposal of bringing the methods of cognitive science to bear on the underlying sources of our beliefs.

Looking beyond the controversial issue of experimental philosophy, it

---

<sup>23</sup> Knobe & Nichols 2008, p. 7.