

Reflections on the Philosophy of Action<sup>1</sup>  
Michael E. Bratman Stanford University  
file: PhilActionFiveQuestions.final.doc - 1/30/08  
For Jesus Aguilar and Andrei A. Buckareff, eds. Philosophy of Action: 5  
Questions (Automatic Press/VIP, forthcoming)  
©2008 Michael E. Bratman

We are both knowers and doers. In each case our minds are appropriately connected to the world. Or so it seems. These commonsense remarks are the background of three central areas of philosophy: epistemology, philosophy of mind, and philosophy of action. In all three cases the issues are conceptual, metaphysical and normative. We seek concepts adequate to an understanding of the phenomena; we seek to understand in what these phenomena consist; and we seek to understand relevant norms – both theoretical and practical – and their relation to matters conceptual and metaphysical. The problems throughout are deep and of great importance in our understanding of who we are; and these problems across these different areas of philosophy interact in complex ways. But I think it is clear that the western philosophical tradition has – with some notable exceptions -- tended to focus more intensely on knowledge and mind than on action and agency. One powerful source of the attraction and excitement of the philosophy of action to me is that the target of its inquiries – our nature as agents – is gripping within self-reflection and of fundamental importance within a wide range of human concerns (some of which I mention below). A second source of the attraction to me of the philosophy of action is the sense that, as compared with our philosophical understanding of knowledge and mind, this is less well understood – though equally fundamental -- territory, and so territory that is more susceptible to our seeing things in new and newly fecund ways.

What is it to be an agent, and not just an object in the natural, causal order? What is it to act? to act intentionally? to act for a reason? What is it for the agent to be the source of the activity, for the activity not just to be the outcome of an event-causal process? What kind of knowledge do we have of our own agency? How is the idea of a normative reason – a reason that is relevant to

the justification of action – connected to the concepts and metaphysics of agency and action? to the phenomenon of motivation of action? to the explanation of action? How is the bearing of rationality on these practical domains – as we say, practical rationality – related to the bearing of rationality on the domains of knowledge and belief – as we say, theoretical rationality? We think that some agents govern their own lives; but what is that? How do these matters look when we turn from the actions and agency of individual human agents to the actions of agency of groups – to our singing the duet, or – to move from small to large – to “we the people”? These, anyway, are some (I do not say all) of the deep, difficult, and complexly inter-related questions that are central to the philosophical study of human agency.

In the background is another question, one that returns to issues about the relation between the practical and the theoretical, to the relation between action and knowledge. The idea of will – roughly, of choice or decision -- seems distinctively relevant to our understanding of agency, and in a way that distinguishes acting from knowing. Choosing or deciding to believe in most cases seems problematic in ways that we begin to get at in our talk of wishful thinking. But choosing or deciding to act and so acting accordingly seems a central case of agency. But what is this idea of will, and what in the world (if anything) does it point to?

One aspect of this idea that is a part of a certain amount of common talk is the thought that a choice or decision in some sense breaks the causal, deterministic chain from the past to the future. I myself find that this “incompatibilist” thought – at least, as a thought about what is essential to choice or decision -- loses its grip when we reflect on how breaking that chain could matter to the metaphysics of the forms of agency we care about. On reflection, what seems to me most important about the will – about choice, decision, and the like – is not whether it is outside the causal order (how would that help us live our lives?), but that these will-like phenomena settle practical questions in ways that impose distinctive structure on our downstream practical thinking and acting. Creatures with what we might as well call “a will” are creatures with the

psychological structures and capacities that are involved in imposing this structure on thought and action.

What structure, and why bother? Here I think both the concept and the phenomenon of planning, and associated norms, are fundamental. Intentions are elements of typically partial and almost-always at least partly future-directed plans. Plans and planning play fundamental coordinating roles in our temporally extended and social lives, especially given our cognitive limitations. That is why we bother with plans. Choice and decision are standard ways we have for arriving at such plan-like states of mind. To understand what the will is we need to understand what it does; and to understand what it does we need to understand how such planning functions in our temporally extended and social agency. Our understanding of that does not depend on seeing the will as outside the natural, causal order. But it does depend on understanding the distinctive roles of planning structures in our practical thinking and action. And that is a matter both of understanding how those structures function, and, relatedly, what the relevant norms (norms, for example, of consistency of intentions and means-end coherence of plans) are that apply to and sometimes guide this functioning. Or so I have over the years argued, in developing what I have called the planning theory of intention and our agency – a theory that can be thought of as a modest theory of the will.

In a way, the planning theory is a response to a debate that didn't quite happen. Philosophy of action as we know it was formed on the crucible of the interactions – primarily in the 1960s -- between Anscombe and Davidson about intention and agency. They agreed about the individuation of action – about the idea that normally when I B by A-ing I perform one action, multiply describable. They famously disagreed about whether the explanation of action by appeal to the agent's reasons for so acting was a kind of causal explanation. And they also disagreed about Anscombe's claim that in acting intentionally one necessarily has a distinctive kind of non-observational knowledge of what one is doing. The disagreement about reasons and causal explanations shaped much of what we know as the philosophy of action. Davidson's enormously influential

desire-belief causal model of intentional agency – a model that sees purposiveness, rather than a distinctive mode of self-knowledge, as at the heart of agency -- is one of the main fruits of this debate. In contrast, the disagreement about self-knowledge in intentional agency stayed for awhile on the back burners, though – due in part to work of David Velleman -- it has lately returned to center stage (together with implications for our understanding of the relation between practical and theoretical rationality). But there was also an underlying agreement – or anyway, a truce – about the very idea of intending to act. For different reasons, neither saw this idea as getting at something fundamental, though in later work Davidson did seek a conservative correction in this skepticism. But in a different tradition of research on agency -- associated in particular with Hector Neri-Castaneda -- the concept and phenomenon of intending to act were seen as fundamental. In particular, intending was seen not just as an output of practical reasoning (which is how Davidson eventually came to think about intending) but as a fundamental element in the inputs to practical reasoning. (A central example is reasoning from intending an end to intending means, though this is not an example highlighted by Castaneda.) The question of how to draw on the insights both of the Davidsonian desire-belief model and of the focus on intending as a basic phenomenon tended to be ignored, though it was brought to the fore in Gilbert Harman's groundbreaking 1976 essay, "Practical Reasoning". The planning theory was my response to this incompletely articulated debate.

And it is a response that has implications for many of the issues noted above that are at the heart of the philosophy of action. One example is that it supports a view of intending as importantly different from believing. Intending  $p$  will play different roles in downstream planning, means-end reasoning, and action than will believing  $p$  (though here I disagree, in different ways, with both Harman and Velleman). As I see it, this also helps us avoid collapsing the norms on intending that are central to our planning agency to theoretical norms, and thereby helps support an important distinction between theoretical and practical rationality. This is tied to recent debates about the nature and ground of

what John Broome calls rational requirements – for example, rational requirements of consistency of belief, and of intention. And one outstanding question here is whether the planning theory can put us in a better position to assess and respond to recent challenges from Joseph Raz and Niko Kolodny to the normative force of such rational requirements.

I also think that the planning theory can make significant contributions to our understanding of basic human capacities for self-governance and sociality – capacities that seem central to what it is to be a human being. I think, first, that when we highlight these planning structures as deeply embedded in the kind of agent we are, we can make progress in understanding what it is for an agent to govern her own life – what self-governance is. This is in part because, or so I think, the possibility of appealing to planning structures helps us answer a deep question that is in the background of our thinking about self-governance: what psychological structures are such that when they guide thought and action the agent directs and governs? (I call this the question of the agential authority of those psychological structures.) And this is in part because of the central roles of planning structures in constituting and supporting major forms of cross-temporal organization of our practical thought and action, and so organization of our temporally extended agency. Appeal to these planning structures gives us a way of understanding the role of the will in self-governance without seeing the will as outside the natural, causal order.

I also think the planning theory helps us better understand the kind of shared agency illustrated by cases like that of our singing the duet, or Margaret Gilbert's example of our walking together – cases of what I call modest sociality. A central problem here is how to understand the distinctive way you and I are related when – to use Gilbert's example -- we walk together, and how this differs from the way in which you and a stranger are related to each other when you are both merely walking down the same street at the same time without bumping into each other. Broadly speaking, the game-theoretic tradition aims to understand our special relation to each other when we walk together as a matter of a kind of equilibrium in a context of common knowledge. But this seems ill-suited to make

the contrast between you and me when we walk together, on the one hand, and, on the other hand, you and the stranger merely walking down the same street at the same time without bumping into each other. An alternative view, associated with Gilbert's work, is to see what is special as certain mutual, reciprocal obligations between us and entitlements of each to hold the other accountable for failing to play her role. In contrast, the planning theory gives us the conceptual, metaphysical, and normative resources to chart a path in the territory between these two approaches: our modest sociality is not just a matter of equilibrium given common knowledge; but it need not (though it usually does) involve mutual obligations and entitlements. Very roughly, what is distinctive of our shared agency are certain structures of interrelated intentions of the participants; and we understand such intentions within the planning theory.

The conjecture, then, is that basic planning capacities are at the bottom of two fundamental human practical capacities: our capacities both for self-governance and for certain forms of sociality. Whether these planning theoretic approaches to self-governance and to modest sociality will be successful is a matter for further research. But it does seem to me at least a potential advantage of the planning theory that it promises to help us characterize important features of human agency that may well turn out to be central to these further basic capacities for self-governance and modest sociality.

An underlying idea is that a central target of our reflections in the philosophy of action are basic structures – basic capacities and modes of thought – characteristic of (what we can call) the will and thereby of our agency. We get at these structures by studying how they function in particular cases of practical thinking and intentional action, and by studying the norms that apply to such functioning. But it is these general structures of agency – structures that help constitute and support our temporally extended and social lives -- that are our primary target. And the conjecture is that these planning structures help us understand other basic human practical capacities – including, in particular, capacities for self-governance and sociality.

Questions about how to understand our agency – about the needed concepts, the underlying metaphysics, and the associated norms – are at the foundation of our understanding of who we are both as individuals and as participants in sociality. Our answers to these questions will impact our approach and answers to a wide range of issues. The concepts and phenomena of agency and intention – both individual and shared -- are at the heart of an enormous range of targets of human reflection. There are countless examples from moral philosophy (there is, of course, the very idea of moral responsibility and accountability; and think also about the distinction between intending and expecting that is built into the principle of double effect), from literature and the arts (think of the ideas of authorial intent, and of artistic collaboration), from social scientific studies (which tend to presuppose certain models of individual agency and practical rationality), from law (think of appeals to intent in the criminal law, and to the very idea of a conspiracy), from artificial intelligence (what exactly are we looking for an artificial version of?), and from efforts to understand the interaction between neuroscience and our self-understanding as agents, both in our commonsense and in domains of accountability, such as the law.

For example, if the planning theory can help us answer the question about agential authority noted earlier -- what psychological structures are such that when they guide the agent governs? – then perhaps we can make better sense of how a neuroscience of mind and behavior, one that traffics solely in an event causal order, can be compatible with the reality of agency, of agents as the source of actions. And we can ask whether this can be part of the effort to understand the compatibility of such a science with fundamental norms of accountability. A related question is whether this helps us better understand current debates, both in philosophy (deriving from Anscombe's work) and in psychology (for example, some of the skeptical views of Daniel Wegner), of the scope and limits of our knowledge of our own agency.

At another extreme, we can ask whether, and to what extent, our understanding of what I have called modest sociality can help us to understand larger forms of social agency. Prominent examples here include the social

activities central to a legal system and/or to democratic politics. Is it fruitful to think of law, and/or democracy, as a distinctive kind of shared activity? (Scott Shapiro has recently been exploring the first question, about law, and Anna Stilz has been exploring the second query, about democracy.)

I am a “let many flowers bloom” sort of philosopher: I doubt we can know in advance which specific research projects will lead to important results. Indeed, in philosophy it is hard to know how significant the results are even once one has arrived at them – these judgments of significance can require years of further reflection on the part of the philosophical community. But it does seem clear to me that our understanding of our agency, both individual and shared, is as fundamental as things get in philosophy. And I am hopeful that a modest theory of the will such as that sketched in the planning theory can help us better understand how to think about our agency and about related matters of practical rationality, self-governance, sociality, and the location of our agency in the natural order.

---

<sup>1</sup> Thanks to Manuel Vargas for helpful comments on an earlier draft.