

FIVE QUESTIONS ON PHILOSOPHY OF ACTION

(FORTHCOMING IN JESÚS AGUILAR AND ANDREI BUCKAREFF, EDS.
PHILOSOPHY OF ACTION: 5 QUESTIONS (AUTOMATIC PRESS/VIP))

MANUEL VARGAS

UNIVERSITY OF SAN FRANCISCO

1. Why were you initially drawn to theorizing about action and agency?

In terms of my own first-personal narrative, the most obvious proximal cause of my theorizing about agency was a graduate seminar on free will taught by Peter van Inwagen. It was my first semester of graduate school, and van Inwagen's forceful presentation of incompatibilism made a big impression on me. I left that course thinking incompatibilism was both obvious and irrefutable. The only problem was that I didn't stay at Notre Dame. I transferred to Stanford in the following year, where I discovered the truth of a remark John Fischer once made: *Indiana is for Incompatibilists and California is for Compatibilists*. Some of the folks there who most influenced me, especially Michael Bratman and Ken Taylor, are thoroughgoing compatibilists. I was really struck by the fact that both of these smart, thoughtful guys seemed genuinely puzzled by the impulse to incompatibilism. I wasn't entirely ready to give up on my incompatibilism (which was by then shifting from libertarianism to hard incompatibilism), but I felt a need to be able to find some way to reconcile it with an appreciation for the appeal that compatibilism clearly seemed to have for some otherwise compelling philosophers. And, so my interest in thinking about free agency and free action began to take root.

So, that's the intellectualized part of the story. But there is also the fact of my local conditions when all of this was going on. For good or ill, I kept taking seminars where the problem of free will would crop up in the course of things. Out of sheer laziness (or, as I like to think about it, out of a dimly sensed need to conserve my energies for later), I kept seizing on the topic as a subject matter for seminar papers, whether the course was on Hume, Aristotle, Nietzsche, philosophy of mind, or philosophy action. Thus, when it came time to write a dissertation, free will seemed like an obvious choice. It was certainly going to require less preparation than any other topic, I (perhaps falsely) thought. So there I went— and here I am.

Having given this first-personal narrative, I feel compelled to flag that I'm actually somewhat suspicious about the status of these kinds of reports. To be sure, accounts like the one I just gave is the sort of response we expect to questions that demand an explanation of what we have done or been drawn to do. But, these first-personal narratives —narratives emphasizing reasons and context-specific practical deliberation— seem susceptible to undermining by a range of alternative explanations. We might, for example, imagine that some day we will be able to give a neurochemical story about why I did what I did. Or, we'll be able to point to a sociological story that explains what I did. Or, we could go in for an old-fashioned deflationary story in terms of unobvious psychological forces of dubious rationality. And, at least on the surface of it, all of these explanations seem to threaten the kind of explanation I initially offered of my attraction to philosophical issues of agency. This isn't to say that I think such explanations have no place, or that action explanations can be reduced or eliminated. I think it is a very difficult matter to say what the relationship is between these various forms of explanation are for human action, and whether one or another should or could rightly displace the sort of first-personal narrative that serves as our default explanatory position. I mention it, though, because these matters reach to the heart of

philosophy of action. Indeed, in some ways it illustrates how the subject matter is ubiquitous and foundational.

2. What do you consider to be your own most important contribution(s) to theorizing about action and agency, and why?

By my lights, estimations of importance of contribution are most fruitfully made by other people near the end of one's career, when further disavowals and obfuscation by the author become difficult (or ideally, impossible). So, I fervently hope that this is a wildly inappropriate time for assessments of my work.

That said, I have no reason to believe that I've made any particularly important contribution to this literature, even by my own lights. Don't get me wrong— I certainly *hope* to make some contribution over the next few decades. And, I am delighted to be included in this volume, as it is chock full of people I greatly admire and from whom I have learned so much. However, I am no peer to the members of this distinguished lineup, so I imagine my inclusion in this volume reflects some optimism about my potential for future contribution more than the judgment that I have already provided a substantive contribution.

Nevertheless, I do have something to say about one way of conceiving of the nature of contributions to philosophy, and how such contributions might be measured. My preferred way of thinking about the nature of philosophical contributions is in terms of an answer to this question: To what extent did the purported contribution help us make progress in understanding the subject matter? Given this metric, there are many ways an account could make a contribution. One way is simple enough— just *be right*. That's a pretty rare thing in philosophy, though. A second way to contribute to progress in understanding an issue is by raising problems, identifying puzzles, or pointing to matters that merit attention. A third kind of progress is neither a matter of being right nor a matter of asking useful questions. Instead, it is a kind of progress that is usually only achieved by being interestingly wrong. It is a familiar enough achievement to make a philosophical contribution by developing an underappreciated possibility and getting other philosophers to respond to it. Of course, the usual result is that we discard the proposal when we understand why that possibility is not the right one. In doing this, though, we can accomplish two things. First, we constrain the scope of remaining viable possibilities. Second, we can trigger new innovations or open up new possibilities that were previously invisible to us. So, I believe that both of these possibilities provide an underappreciated route by which a philosopher can make a contribution to understanding some subject matter.

One consequence of this picture is that it is misguided to conceive of philosophical progress (and philosophical success or failure) in the paradigmatic case as the achievement or failure of some lone, heroic figure. Instead, I think we are better off thinking about these things in terms of the success or failure of a collective, knowledge-seeking enterprise that (hopefully) makes progress on the backs of our collective work and interaction. So, even if no one individual has got the right account of intentional action, free will, or what have you, we might still be collectively making progress towards achieving the True Account. I do not wish to deny that there might well be individuals who have made considerably greater contributions in the course of things. But, it is important to remember that those contributions are also collective, requiring large networks of education and idea dissemination, as well as a cultural context that facilitates and takes seriously the kind of inquiry that yields some philosophical achievement.

It is also worth noting that even in the case of an individual, it is quite likely that we'll miss the mark in our assessments of his or her work. Here's why: we don't normally track actual contributions in the sense I have suggested. It could turn out that whatever contribution a given philosopher makes is not appreciated for the role it *in fact* plays in our collective project of truth seeking. Being demonstrably wrong about some matter, and forcing the literature to point it out (or inspiring someone with the right professional visibility to highlight his or her avoidance of the error) can permit us to collectively move on to some other matter. In doing so, it can sometimes be the crucial thing needed to move us closer to the right account. Note, however, that it is also exactly the sort of thing we tend to ignore or fail to recognize in our assessments of philosophical contributions. In philosophy, instructive failure is oftentimes more fruitful than (let's admit it, a usually short lived) "success." Yet, we rarely celebrate such failure and we virtually never measure accomplishment by it. Moreover, even when we attend to useful failures, the utility of the failure is not always obvious to us. So, although I think there is a reasonable metric for thinking about what constitutes a greater and lesser contribution to philosophy, I'm also skeptical about whether we're very often in a position to accurately evaluate the matter.

In the spirit of answering the likely intent of the question, though, I will say something that more directly addresses the substance of the question. If I were to guess about what other philosophers are likely to cite as my contributions to the field thus far, I'd point to two things. One concerns my approach to the problem of free will and moral responsibility, something I call *moderate revisionism* (Fischer, Kane, Pereboom, & Vargas, 2007; Vargas, 2004; Vargas, 2005a). This approach is a response to the thought that our commonsense conceptions of free will and moral responsibility cannot be made consistent with an independently plausible, broadly naturalistic picture of the world. (See my answer to the above question about why I got bugged by this issue.) Moderate revisionism is a way of responding to the problem of reconciling our self-image with a naturalistic picture of the world, which I take to be independently plausible. The answer I have been developing is one that abandons some particularly problematic elements of commonsense, and attempts to show how we can re-anchor our understanding of free will and moral responsibility in things that do not depend on our being agents of the sort described by libertarianism.

If this view has any contribution to make, in the sense in which I think of philosophical contributions, it might be the following. First, it helps bring to the foreground the role of intuitions in our theory building in this domain. Consequently, it makes particularly salient the way in which philosophers have tended to build their metaphysical commitments out of commonsense. Secondly, it helps to sharpen questions about the relationship of the normative to the metaphysical. As I see things, anyway, this is a domain in which we can make excellent progress if we put aside our pre-philosophical convictions and instead ask ourselves what conception of agency is required to justify our practices of praise and blame, and what is needed to license the judgments we make of these things.

The second philosophical contribution my work might be taken to have concerns the articulation of a kind of puzzle raised by the relationship of an epistemic requirement on responsibility and the conditions under which we acquire various capacities that are taken to constitute our free and responsible agency. The epistemic requirement is, roughly, the requirement that an agent needs some awareness of likely outcomes of the action in order to be held responsible for it. In "The Trouble With Tracing" (Vargas, 2005b) I argue that either the knowledge condition isn't satisfied when we acquire a significant number of our action-determining capacities and dispositions, or else we have an inadequate account of the knowledge condition.

I am not the only one who has picked up on this puzzle, but I think my discussion of it has provided a useful target for some further philosophical rumination on the matter. It remains to be seen whether the issue has any legs, though. Moreover, I think it is an open question whether or not the inevitable dissolution of the problem moves us closer to the truth about the conditions for moral responsibility.

3. What other sub-disciplines in philosophy and non-philosophical disciplines stand to benefit the most from philosophical work on the nature of action and agency, and how might such engagement be accomplished?

Psychologists and neuroscientists interested in agency and action have a good deal to learn from philosophical work on agency. And, there is plenty philosophers of agency can learn by studying work in psychology and neuroscience. Still, even a cursory glance at work by scientists interested in agency too often reveals impoverished conceptual resources when it comes to interpreting their own data or drawing out the philosophical ramification of the work. There are, now, more philosophically minded folks who have started to wade into these matters, but there is a long way to go.

Inside philosophy, it is less clear to me how much philosophers in other subfields might learn from studying the philosophy of action, at least those parts with which my own work is concerned. My doubt is partly driven by the thought that philosophy of agency — a term I use to apply to both philosophy of action as it is traditionally conceived of as well as more general issues of agency that include things like free will, moral responsibility, and autonomy— appropriates so much from neighboring fields. Philosophy of agency integrates issues in metaphysics, philosophy of mind, moral psychology, and ethics. In this, it is like many other fields in philosophy, but more so. That is, philosophy of agency integrates issues from a wider range of fields than most fields in philosophy, and consequently, we are still absorbing much of what we need know. Until we've absorbed a bit more, it isn't clear to me that we're going to be offering much to other areas in philosophy. Naturally, I hope I am wrong. And, I think, there are pockets of philosophy agency where I am clearly wrong. However, within those areas I focus on, where there is some direction of influence running the other way (from philosophy of agency to some other subfield), the influence tends to come from the roughly normative aspects of work on agency. And, it seems to me that perhaps the most promising direction from work on agency to normative issues that ought to be pursued by someone— perhaps an ambitious graduate student?— involves connections with political philosophy. There has been some work on notions of autonomy in political liberalism that have grown out of work that was done in the context of work in philosophy of action. And, Sam Scheffler has done interesting work on the way in which conceptions of agency and responsibility interact with various issues in political philosophy. However, my sense is that there is more to be said on these matters.

4. What do you regard as the most neglected issues in contemporary work on action and agency that deserve more attention?

I can't speak with authority about philosophy of agency more generally, but there are several issues connected with work on free will and moral responsibility that seem to me to be neglected. For example, philosophical accounts of moral responsibility are not particular useful at offering guidance in real world circumstances. Whether or not it is a vice that our theories do not ordinarily

offer useful guidance in real world contexts, I think such a service would be an undeniable virtue. However, I am inclined to think that there are at least two barriers to the development of action-guiding theories of moral responsibility.

First, we do not have a good grasp of the ways in which situations structure the powers of agents. Philosophers working on agency have tended to think about the powers of agents in atomistic terms. That is, philosophers have tended to think of agents as self-contained things to be understood entirely detached from a context or environment (including psychological and cultural). Such pictures of our agency are, I think, deeply flawed, or at the very least, profoundly misleading. If our powers are partly structured by our environments, until we have a good understanding of the ways in which this interaction between agent and environment occur, we will not be able to provide much guidance in real world cases— precisely because real world cases *are* cases embedded in environments.

A second barrier to the attainment of action-guiding theories is that we lack an epistemology of responsibility in non-ideal circumstances. That is, we do not have a philosophical account of how to make judgments of responsibility given the messiness of real-world circumstances, circumstances where full information about the agency of others and the deliberative circumstances of their choices is impossible to secure. Moreover, growing scientific skepticism about the veracity of even well intentioned first personal reports is surely no help. Given the consequences of moral praise and blame, what is needed is some account of how we might get reasonable evidence about the powers of agents as they are relevant to responsibility, along with some standard of deciding what counts as adequate evidence given our epistemically imperfect circumstances. This is not to say that we do not make such judgments all the time. And, indeed, legal assessments of responsibility have grappled with a version of this problem for a long time. However, it seems to me that in the case of philosophical accounts of moral responsibility we have not even begun to do this work.

Two other problems strike me as deserving of more attention: the epistemic condition on moral responsibility (roughly, the idea that agents have to know something about the consequences of their action in order to be appropriately held responsible) and the relationship of risk to blameworthiness. I do not have any sense of how these matters should be sorted out, but I do think it would be worthwhile for philosophers interested in responsibility to think more about these issues.

5. What are the most important open problems in philosophical theorizing about action and agency, and what are the prospects for progress?

When it comes to important open problems connected to action and agency, we have an embarrassment of riches. Among them, I'd say that some of the most important and difficult problems concern the connection of the normative to the natural, and the matter of how we account for the ontology of the normative dimensions of human agency (including rational, epistemic, and moral aspects). I can imagine some philosophers disputing whether these matters are properly in the domain of philosophy of action, but even if they are not I think they are clearly in the domain of philosophy of agency.

The ubiquity of the problem I mentioned —the intersection of the natural and the normative in human agency— can be seen throughout the existing literature. We see it in attempts to understand the nature of practical reason and the integration of agent-based and world-based inputs to it; we see it in debates about the causal theory of action and the place it gives to reasons

in a world of causes; we see it in disputes over the conditions of free will and responsible agency; and, we see it in discussions about the role of knowledge as it conditions and structures various forms of agency.

I am also inclined to think there are a number of open problems connected to the literature of free will that we are just starting to address in a direct and fruitful way. So, for example, I think we are on the verge of small cottage industry surrounding the matter of desert, including whether and how it is relevant to the success or failure of various accounts of free will and moral responsibility. I also think we are finally engaging with some deep and complex methodological issues surrounding how we build accounts of free will and moral responsibility. The matter of intuitions and their relationship to the metaphysics of free will seems to me to be a crucial and unresolved issue.

I also believe there is a growing sense that the dominant jargon of the field is oftentimes as much hindrance as help. For example, the fixation on the compatibility-with-determinism debate, and the attendant emphasis on whether one is a compatibilist or incompatibilist can sometimes obscure threats that have little or no direct relationship to determinism. My claim is not that we cannot usefully deploy the language of compatibility. Rather, my point is that given the varieties of matters around which compatibility and incompatibility talk arise (including free will, moral responsibility, deliberation, maximally desirable forms of agency, etc.) the language of compatibilism and incompatibilism *simpliciter* has a utility comparable to the distinction between realism and antirealism in metaphysics, or as internalism and externalism in moral psychology (which is to say: very little).

I don't have a confident assessment about the near-term prospects for progress on these matters. However, I do remain optimistic about the possibility of our collective progress, over larger lengths of time. Even if the flood of false theories spilling from the lips of philosophers doesn't abate any time soon, identifying *why* they are false is an important kind of progress available to us. And, I think prospects are quite good that we will some day come to understand why nearly all of our going theories, mine included, are on these matters, mightily mistaken. I confess to being cheered by this possibility.

BIBLIOGRAPHY

- Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M. (2007). *Four Views on Free Will*. Malden, MA: Blackwell.
- Vargas, M. (2004). Responsibility and the Aims of Theory: Strawson and Revisionism. *Pacific Philosophical Quarterly*, 85(2), 218-241.
- Vargas, M. (2005a). The Revisionist's Guide to Responsibility. *Philosophical Studies*, 125(3), 399-429.
- Vargas, M. (2005b). The Trouble With Tracing. *Midwest Studies in Philosophy*, 29(1), 269-291.