

Building A Better Beast

Manuel Vargas

1. Agents in a world of causes

Here is a worry that sometimes occurs to people long before they get exposed to professional philosophers: if everything in the world is caused, and I'm just another part of the world, how is it that I'm the one really doing anything? How can I be said to do things of my own free will? How can I be blamed for anything? With enough schooling of a particular sort, the worry is sometimes refined to something like this: Given the truth of the causal closure of the physical, everything we do is a causal upshot of prior causes or perhaps more basic forces. If so, then how are there some causes (that *you* did it, or that *I* did it) that merit a distinctive sort of moral evaluation while there are other causes (the storm destroyed the house, the infant did it, it was a scared puppy, etc.) that do not merit this distinctive moral evaluation? Is there a principled way to draw this distinction, given that they are all causes?

The first thing to say in reply is that not all causes are equal. Motives and context matter. A declaration of love will be caused by human motivations, but the who, when, and why of someone saying "I love you" makes a big difference to us. The issue isn't whether we can distinguish between causes, or whether causes can have different valences to us. Rather, the issue is to say (1) what those features are that mark us off from the parts of the universe that we do not praise and blame in a distinctively moral way, and (2) why it should be those features and not some other in virtue of which we make this distinction. As a first step towards answering these questions, it is useful to begin by reflecting on what is required to appropriately hold someone morally responsible. Answering this question may not yield a complete answer to the question of what is required for any moral evaluation (for example, we may find that something might be evil without being morally

responsible). Nevertheless, a good account of moral responsibility would show how large swaths of our moral evaluations can be appropriate in a causally ordered, maybe physicalist world.

In this paper I will argue for a particular account of morally responsible agency and contrast it with alternative accounts. My chief aim in this paper is to motivate a kind of view about responsible agency, and to say how it is similar to and different from other views. This sort of project is related to the rich array of issues that have been discussed under the heading of ‘free will’. However, much of what is central to those discussions—e.g., debates about the compatibility of free will with determinism—is mostly peripheral to my project. I do think incompatibilism is true as a description of common sense, but that there are various reasons to favor a recasting of our concepts in ways that do not invoke incompatibilist commitments. My account can be characterized as moderately revisionist. If it turns out that the correct considered view we *should* have about responsible agency is incompatibilist, then its main import here would be to add a further condition to my account of responsible agency.

2. On the road to BASRA

It is useful to distinguish between a theory of moral responsibility norms and a theory of responsible agency. The latter is concerned with what candidates are subject to responsibility norms. The former is concerned with what those norms are. Carving up the terrain in this way runs the risk of misrepresenting the degree of independence that exists between a theory of responsible agency and a theory of moral responsibility norms. Although it might be possible to specify each in the absence of the other, and while we might imagine a wide variety of combinations between the various species of theories of responsible agency and theories of responsibility norms, these two classes of theories can and should compliment one another. In this paper I want to focus on the responsible agency part of

the equation, but in order to do so I will at various points in time appeal to considerations grounded in a theory of responsibility norms.¹

As I see it, it is helpful to think of our responsibility norms and practices as, roughly, a subset of moral norms and practices, directed at building better agents, agents that appropriately respond to moral considerations. At its best, the web of responsibility-characteristic practices, attitudes, and judgments, get creatures like us — beasts of a peculiar sort — to, over time, better respond to what moral considerations there are. The system isn't perfect. Circumstances can undermine the point of various normative practices. Agents can, through accident, context, or birth, come to be immune to the normal mechanisms by which the responsibility system operates. Nevertheless, an adequate system of moral responsibility is one that is generally effective at getting creatures like us to, over time, become better at appropriately responding to what moral considerations there are.

My principle concern here is to explore some dimensions of responsible agency that mark us out as the kinds of things to which the norms of responsibility properly apply. The task of a theory of responsible agency is to characterize what I will call the Basic Structure of Responsible Agency (BASRA), a model of the features that are at least jointly sufficient for responsible agency in the ordinary context.² The basic elements of agency must make it in, for responsible agency is not possible without bare agency. Presumably, these basic capacities involve things like instrumental reasoning, the capacity to desire, to believe, and to intend. We should also suppose that an agent has

¹ I have attempted to develop such an account, at least in outline in Manuel Vargas, "Moral Influence, Moral Responsibility," (Unpublished manuscript).

² In an earlier paper I called this "The Basic Agential Structure of Responsibility" which yielded the absolutely horrible acronym BASR. BASRA is still horrible, but not absolutely so, as it at least has the virtue of being pronounceable. So, I've elected to revise the name and acronym I used in Manuel Vargas, "On the Importance of History for Responsible Agency," *Philosophical Studies* (forthcoming).

reasonably reliable capacity to foresee the consequences of his or her action.³ This much is non-controversial and can be shared by a wide range of accounts of BASRA.

What is more contentious is the further claim that BASRA requires the capacity to recognize or detect moral considerations in the relevant circumstance and a capacity to appropriately govern one's conduct in light of them.⁴ On the account I favor, this is the core or most salient feature of responsible agency.⁵ To be sure, sensitivity to moral considerations need not be conscious, and the agent need not recognize that it is a moral consideration *qua* moral consideration that is moving him or her to act. What does matter is that it is a moral consideration that moves the agent to act. How and whether we judge the springs of our own action is less important than what in fact moves us to act.

Two comments concerning moral considerations are in order.

First, moral considerations are immensely varied. Roughly, moral considerations are anything relevant to agents guiding their conduct in light of morality's demands. At a very general level, what this amounts to is determined by (1) the demands morality imposes on us, (2) the contexts of action, and (3) the kinds of agents with which morality is concerned. The extent that a particular moral consideration is a consideration for *me* will depend on what morality requires of me, what

³ This is sometimes characterized as the "knowledge condition" on moral responsibility. In my view, this very general feature of agency should be distinguished from the sensitivity requirement I discuss below. My thanks to Michael McKenna and Kevin Timpe for pressing me to clarify this point.

⁴ In broad outlines, views similar to this have been held by a number of figures. Indeed, at a sufficiently broad level of description this sort of view may well be the most prevalent general approach to responsible agency, at least among compatibilists about moral responsibility and determinism. What unifies these accounts, to the extent to which they have any unity, is a consensus about the importance of reasoning capacities for responsible agency. The details, of course, differ significantly. Views of the sort I have in mind can be found in, for example: Nomy Arpaly, *Unprincipled Virtue* (New York: Oxford, 2003), John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998), Michael McKenna, "The Limits of Evil and the Role of Moral Address," *Journal of Ethics* 2, no. 2 (1998), R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1994), Susan Wolf, *Freedom within Reason* (New York: Oxford University Press, 1990). My view has some affinity to the views of Fischer and Ravizza and Jay Wallace, although we differ about a number of important aspects of the theory, as will become evident by the end of § 4.

⁵ I am officially agnostic about the complex issue of whether some identification condition needs to be included in BASRA. If it does, however, I think it will have to be in addition to the moral considerations-dependent conditions I am here concerned to describe.

circumstances I find myself in, and facts about what it takes for me to guide my conduct in light of morality's demands.

Second if we were to have in hand a worked out account of what a moral consideration consists in, I strongly doubt that we would be able to tell a single, unified, and true story about how sensitivity to moral considerations works across all contexts. Sensitivity to moral considerations is unlikely to be a unified phenomenon. The considerations of morality are simply too diverse in their content and in their manifestations for us to be sensitive to them in any single way. Consider that some moral considerations may take the form of reasons understood in some robust sense, for example, something involving belief-like propositional content that could be articulated. Other cases of sensitivity to moral considerations may involve some complex process of visual recognition. Being able to perceive when a quiet friend is in need of consolation is unlikely to involve the same mechanisms as those involved in the exchange of reasoned philosophical arguments about duties. And, at other times, sensitivity to moral considerations may involve susceptibility to a nagging feeling, reacting to a dim hope, being able to imagine the situation of another, or attending to an inarticulate, largely inchoate suspicion about things. Just how this happens may vary from person to person and from situation to situation. What capacities are involved in recognizing moral considerations in a given context might vary with the features of the context. The power of a particular agent's detection capacities in light of situational factors might vary considerably across contexts.⁶ And, there might well be people or creatures whose sensitivity to moral considerations

⁶ Although the significance of situationist social psychology for ethics remains controversial, I take it that one of the clearest ways in which it is relevant is in suggesting that social contexts can have powerful effects on our perception of moral considerations. For recent philosophical work on these issues, see John Doris and Dominic Murphy, "War Crimes," (Unpublished manuscript, 2004), John Doris and Stephen Stich, "As a Matter of Fact: Empirical Perspectives on Ethics," in *The Oxford Handbook of Contemporary Philosophy*, ed. Frank Jackson and Michael Smith (Oxford: Oxford, forthcoming), Dana Nelkin, "Freedom, Responsibility, and the Challenge of Situationism," *Midwest Studies in Philosophy* 29, no. 1 (2005). A less sophisticated but somewhat more familiar concern has to do with the way in which aspects of contemporary society systematically degrade moral sensitivities. Depending on one's convictions the targets of concern might include video games, music, and movies, as well as wealth, privilege, race, and gender.

relies on very different underlying psychological mechanisms than the ones typically relied upon by normal human adults.⁷

Even in the case of articulated moral reasons of the sort moral philosophers endeavor to trade in, aspirations for a clear and unified account of moral considerations understood in this fashion seem dim. First, we do not have anything like consensus about what constitutes a moral reason. Second, we do not yet have a good grasp of the psychological mechanisms involved in recognizing and governing our behavior in light of reasons in general, much less specifically moral reasons. The connection between one's detection of moral considerations and moral motivation is something about which there remains significant philosophical dispute. Despite these barriers, I trust that the idea of sensitivity to moral considerations is reasonably intuitive, even if the mechanisms by which it occurs are hardly obvious. (In this respect, I might be somewhat similar to consciousness—familiar but exceedingly difficult to characterize.)

3. Capacities and Context, or I am Myself and My Circumstances

Let us return to the idea that responsible agency consists in two elements: a capacity to be sensitive to moral considerations and a capacity for self-governance in light of those considerations. I'll begin with some general remarks about the language of capacity in this context, and move to more specific considerations about sensitivity to moral considerations and self-governance.

We should acknowledge that there are perfectly legitimate uses of 'capacity' whereby one can have a capacity and fail to act in a way consistent with the full or best use of that capacity. Although I am not speaking in Spanish at the moment, there is a clear sense in which I retain the capacity to speak Spanish. And, although I overcooked the fish, I had and continue to have the capacity to not overcook it (indeed, I have undercooked it a number of times!). What we have in mind with these

⁷ Autistics may be an interesting case along these lines. See, for example, Jeanette Kennett, "Autism, Empathy and Moral Agency," *Philosophical Quarterly* 52, no. 208 (2002).

sorts of capacity claims varies. Sometimes all we mean is something like “given what I know about the case, the following outcome or outcomes are consistent with it.” Other times we might mean capacity talk in something like the following way: “in cases relevantly similar to this one, things tend to go this way.” And, oftentimes, talk of capacities is meant to describe some power one might have and can use under a wide range of conditions but not all conditions. For example, even when I’m sleeping, I still retain my capacity to speak Spanish. I simply can’t exercise it until I am awake. It would seem really strange to insist that I lose my capacity to speak Spanish every night and regain it every morning (Where did it go in the intervening time? How did I relearn the Spanish language so quickly?). So, at least sometimes, when we speak of capacities we do so while acknowledging one might have a capacity without exercising it or even having an opportunity to exercise it.

Here is where concerns about the metaphysics of free will intrude. Even if it is reasonably clear that there are a variety of ways in which something might truly be said to have a capacity, we might wonder whether there is a considerably more demanding notion of capacity required for ascriptions of moral responsibility. Consider a version of capacity talk that I will call the Garden of Forking Paths (GFP) picture of capacities. On the GFP model I have in mind, the responsibility-relevant notion of capacity is very demanding, requiring that we hold fixed *all* the background conditions.⁸ It can be illustrated by asking whether you can walk through a garden via a different footpath than the one you are on. It all depends on the garden. In some gardens, the path branches with lots of ways to walk through the garden. In other gardens, there is only a single path. On the GFP model, capacity talk is construed along the lines of being in a garden with forking paths. In a deterministic universe, there is only one path, and thus, the only GFP capacities you have are those

⁸ Two points: First, I do not mean to suggest there is no room for alternative possibilities-requiring compatibilism. I simply want to make use of a model with stricter requirements. Second, I believe it was John Martin Fischer who introduced the GFP metaphor to the philosophical literature. His usage of it derives from the wonderful Jorge Luis Borges story “The Garden of Forking Paths” which can be found in Jorge Luis Borges, *Collected Fictions* (New York: Penguin, 1998).

you exercise. There are no unexercised GFP-type capacities. In an indeterministic universe there are multiple paths, and thus, unexercised capacities.

I am not aware of any reason to suppose that our talk of capacities is uniform across all contexts of action. In some contexts, we might well mean capacities in the GFP way of putting things. In many everyday contexts I suspect we do not require a GFP sense of capacity. In the course of a conversation we might put pressure on one or another way of putting things, dubbing a favored sense the “real” sense of capacity. However, these locutions are akin to instances of distinguishing between the beautiful and the *really* beautiful, where the italicized ‘really’ merely marks satisfaction of some more demanding conception of beauty. Still, there are senses of capacity that are customary for some notions and not others. In the context of moral responsibility, however, it will come as no surprise that incompatibilists typically hold out for a very strong notion of capacity (GFP, for example) and compatibilists typically hold out for a weaker notion.

On the *prescriptive* (and likely revisionist) account I offer, the notions of capacity invoked in the detection and self-governance requirements of responsible agency should be read as something akin to the capacity to speak Spanish, and *not* as a capacity of the GFP sort. One can retain capacities of this sort without having the opportunity to exercise them, and one can retain these capacities even if we were to learn that the world was and is deterministic. (Try it: even if the world has been deterministic for its entire history, there is surely *some* sense in which I retain the capacity to speak Spanish even when I am not speaking Spanish). On this account of capacity, one can have a capacity to detect moral considerations and to appropriately regulate one’s behavior in light of them without actually having done so.

This said, I am friendly to the idea that at least with respect to a capacity for volitional control, the GFP model or something very much like it describes how many, perhaps most of us tend to think about a range of issues concerning human agency, including free will and moral

responsibility. It is an idea that many of us have of ourselves, at least sometimes. Moreover, what empirical data there is about these things seems to support this conclusion.⁹ However, for a variety of reasons I do not find this picture of ourselves to be especially promising, as matter of both broadly scientific and philosophical plausibility. Crucially, I do not think the kind of agency required to make sense of the GFP model is required for moral responsibility, even if the GFP model is how we tend to think about things. The right account of responsible agency might be different than we suppose, and this goes for the right notion of capacity required for our self-governance in light of moral considerations.

As we have seen, there is a range of capacity notions, from strong to weak, and I am dubious that there are adequate considerations for why—apart from the way we may happen to think of things — we *should* think of the capacity for self-governance in an especially demanding way. To my mind, the project we should be engaged in is the construction of the best account we can offer of responsible agency that is (1) compatible with our best knowledge about the kinds of creatures we are and (2) that provides a justification for the normative work we demand of a notion of responsible agency. I suspect that our best account of responsible agency will very likely depart in principled ways from the way our cultural inheritance (informed by substance dualism and other philosophically disreputable views) may have taught us. Progress in any domain oftentimes requires the abandonment of prior, mistaken views. The situation is no different in the philosophy of moral agency.

⁹ There has been some interesting evidence that suggests that common sense is *not* incompatibilist. See Eddy Nahmias et al., “Is Incompatibilism Intuitive?” *Philosophy and Phenomenological Research* (Forthcoming). However, for responses and other pieces of evidence that our folk concept is incompatibilist, see Shaun Nichols, “Folk Intuitions on Free Will,” *Journal of Cognition and Culture* 6, no. 1 & 2 (Forthcoming in 2006), Shaun Nichols, “The Folk Psychology of Free Will: Fits and Starts,” *Mind and Language* 19, no. 5 (2004), Shaun Nichols and Joshua Knobe, “Moral Responsibility and Determinism: Empirical Investigations of Folk Concepts,” *Unpublished manuscript* (2004). I have discussed some of this work in Manuel Vargas, “Philosophy and the Folk: On Some Implications of Experimental Work for Philosophical Debates on Free Will,” *Journal of Cognition and Culture* 6, no. 1 & 2 (forthcoming in 2006). That said, the literature is still unfolding.

So, the notions of capacity at stake in the detection and self-governance requirements of responsible agency are ones that do not require a GFP model of capacity. Specification of exactly what the relevant notion of capacity amounts to is a tricky business, and for the moment I will leave things sketchy. I will return to these details in a bit, after consider the first of the two requirements I mentioned: the capacity for perceiving moral considerations.

The importance of a capacity for sensitivity to moral considerations becomes clear once we consider what things would be like to hold responsible an agent that altogether lacked the capacity to perceive moral considerations. Such a creature would hardly be an appropriate target for a system of norms, practices, and judgments directed at improving sensitivity to moral considerations. However, we might wonder about cases where an agent does not altogether lack a capacity for detection but lacks it only intermittently, or in particular contexts. If an agent is not culpable for having worked him or herself into this circumstance, then the agent should not count as a responsible agent in that particular instance or circumstance; it would not be reasonable to subject such an agent to the norms of responsibility. It is entirely plausible that an agent might culpably work him or herself into a state where that agent lacks the capacity to respond to moral considerations. What then?

To address this concern, it is useful to introduce some conceptual machinery. A structural account of morally responsible agency is one that holds that whenever the BASRA is present, you have a responsible agent. When it is absent, the agent is not responsible.¹⁰ End of story. A semi-structural account adds a further twist. On a semi-structural account, historical considerations come into play in a very specific way. When BASRA is absent, it matters why. If the agent never had BASRA, the agent is not the sort of thing to which the norms of responsibility apply. However, if the agent did have BASRA, but knowingly and intentionally took steps to eliminate it, then the

¹⁰ Harry Frankfurt's early work is perhaps closest to a pure structural theory. See Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, no. 1 (1971).

norms of responsibility do apply, in light of the agent's current state being something for which the agent is morally responsible. On a semi-structural account of responsible agency, history matters whenever (and only when) BASRA is absent. Structural and semi-structural accounts of responsible agency contrast to historical accounts; the latter include a historical condition irrespective of whether or not BASRA is present.

To get a sense of how semi-structural accounts work, consider the following example. A local platoon commander has ordered his soldiers to round up twenty innocent natives. After they are rounded up, he asks a visiting guest to shoot one, and if he does so, he will release the other nineteen. When the guest declines, the commander turns to two of his best aides and asks them to kill the twenty natives by the next morning. One aide, Bernard, decides he has to get drunk to be able to follow the commander's order to slaughter the defenseless natives. Only then will he be sufficiently numb to the horror of it all to go through with it. He eventually succeeds in getting sufficiently drunk such that he has destroyed his BASRA. He goes on to carry out the orders, killing half of the natives. The other aide, Phineas had the misfortune of suffering an unexpected head injury some time prior to receiving the order to kill the natives. The head injury results in brain damage of a sort that destroys some of the capacities of his BASRA (he is no longer sensitive to moral considerations, at least in contexts like these), but it has left his other faculties intact.¹¹ Phineas straightaway fulfills his assignment, killing half of the natives. On a semi-structural view of things, Bernard is responsible. In contrast, his colleague Phineas is not. Phineas lacks BASRA, but lacks it through no fault of his own. Therefore, he cannot be appropriately held to the norms of

¹¹ Phineas is very loosely modeled on the real world example of Phineas Gage. The case of Phineas Gage is engagingly described in Oliver W. Sacks, *The Man Who Mistook His Wife for a Hat and Other Clinical Tales* (New York: Summit Books, 1985).

responsibility. There might be good reason to quarantine him, but genuine moral praise and blame have become inappropriate.¹²

My account of responsible agency is semi-structuralist. Since you can only be culpable if you were a responsible agent in the first place, and since the presence of BASRA just is the mark of responsible agency, culpable insensitivity to moral considerations would require that one took some steps to render one immune to moral considerations — i.e., one took steps to eliminate BASRA. In other words, we can characterize the detection of moral considerations without appealing to a robust notion of capacity. We can make sense of the idea that one *should have* been sensitive to those moral considerations, even if he/she wasn't in the actual sequence by asking if the agent intentionally and knowingly did something to undermine his or her ability to detect moral considerations in those circumstances. If the answer is yes, the prior negligence grounds culpability. If the answer is no, the agent is not appropriately subject to the responsibility norms.

So, on this account, we can acknowledge that capacity to detect moral considerations is ordinarily required for responsible agency, but that when that capacity is absent an agent might yet be culpable if he or she knowingly brought it about that the capacity was made absent later on. To be sure, there are further complexities here when it comes to assessments of blame and what an agent is to blame for. Consider, for example, an agent that knowingly undermines his or her capacity to detect moral considerations with an eye towards enabling immoral behaviors, as Bernard did (or as presumably some high school and college students do on a given Friday night). It may make sense to treat this sort of case differently than a case in which someone knowingly takes steps that undermines his or her capacity for sensitivity to moral considerations as, say, a release from the pressures of daily life but without an eye towards enabling downstream mischief. These complexities

¹² Nothing I have said here argues against a more thoroughgoing historical account. I have, however, argued against these accounts elsewhere. See Vargas “On the Importance of History for Responsible Agency.”

are more appropriately left for an account of the particular norms of responsibility, however, and are not necessary to flesh out an account of responsible agency.

Let us now consider the second element of this account of responsible agency, the capacity to appropriately govern one's conduct in light of moral considerations. Here, the notion of volitional control becomes central. Volitional control is a capacity whose exercise or failure of exercise is what renders an agent capable of detecting moral considerations into an appropriate target of moral evaluation. Here is an initial, albeit provisional, proposal for how to construe the relevant notion of capacity:

An agent has the capacity to govern him or herself in light of moral considerations in the relevant context if, either the agent is in fact motivated to act in a way that appropriately reflects those moral considerations, or, were that agent to have been prospectively subject to an appropriate prospective judgment of praise or blame for the considered action or outcome, he or she would have acquired some motivation to act in a way that appropriately reflects those moral considerations.¹³

On this construal, there is no requirement that the agent would have successfully acted in the appropriate way. Instead, what is required is only that the agent has, or would have acquired motivations of the right sort. This construal permits the possibility of weakness of will and recognizes that the considerable power of one's non-moral (or even anti-moral) motivations is not, by itself, enough to disable one's capacity for self-governance in the sense required for moral responsibility.

¹³ To be sure, there is more finessing that needs to be done here to address perverse cases, such as an agent that has the unusual condition of losing the motivation to act in a way that reflect moral considerations only under circumstances where he or she is prospectively subject to blame. My sense is that marginal cases such as these will be rare or maybe even non-actual, so they don't weight against the practical utility of this conception of capacity for a revisionist account of responsible agency. Still, it would be useful to have a principled way to rule out perverse cases that might fail to be counterfactually motivated but that nonetheless possess a capacity for self-government appropriately tied to a system of practices and judgments concerned with moral considerations in the way I have described.

This account does owe something to the traditional conditional analysis of ‘can’ (where ‘could’ means, roughly, one would have if one had wanted to). The difficulties of the conditional analysis are well known, and I will not try to canvass them in full detail here. Instead, I will make three points to make about why the approach I favor has some appeal in this dialectical context.

First, some of the chief problems of the conditional analysis are not problems for my appeal to capacities of both detection and self-regulation, given that my account is intended as a potentially revisionist account of responsibility. The classical conditional analysis faces some devastating objections about the way its analysis of ‘can’ fails to capture the way we use the word. Inasmuch as the conditional analysis was presented as a piece of conceptual analysis about ‘can’, this was a serious problem. Objections of this sort are not a problem for my account, given that it does not endeavor to provide an analysis of our common or ordinary analysis of the responsibility-relevant notion of ‘can’ as it pertains to self-governance.

The second point is that the notions of capacity in detection and self-regulation are, by design, responsive to several philosophical pressures, including an account of the aim of a concept of responsibility, the normative work we expect such a concept to do, and the constraints of providing a plausible characterization of responsible agency. The revisionist cannot stipulate his or her way to freedom, as it were, and the account I offer is not *ad hoc*. On the account of the aim of a system of moral responsibility I favor, the capacities of self-governance relevant to responsible agency should be thought of as tied to a network of practices directed at fostering our sensitivity to moral considerations, given our psychologies and the circumstances we find ourselves in. Conditions that disrupt our sensitivity to detect and/or to appropriately govern ourselves in light of moral considerations undermine our responsible agency precisely because they turn us into the sorts of things for which norms of responsibility do not make sense. Consider, for example, a mental disorder. In some contexts it might have no impact on the suitability of an agent for the demands of

moral responsibility. When it comes to helping someone injured, the presence or absence of Tourette's Syndrome will typically have no significance for the appropriateness of the demands of morality. In different contexts, Tourette's would have no significance for the capacity to govern oneself in light of moral considerations to aid. In other contexts Tourette's might have very strong significance for the capacity to govern one's conduct in light of moral considerations. It might, for example, make it virtually impossible for the agent to preface a question with the standard niceties during the question period following a mediocre philosophy talk.

(Something similar can be said of the capacity for detecting moral considerations. The relevant notion of capacity will be one tied to aspects of what kinds of capacities, over time, and for creatures like us, tend to be those that foster sensitivity to moral considerations. These will be capacities that are tied to somewhat fine-grained similarities of deliberative context, to be sure, but not so fine-grained as GFP-like capacities would suggest, for reasons I've already alluded to when I rejected a GFP-like construal of capacity-talk.)

A third feature of my account is that it recognizes the dubious of supposing that the kind of self-governance required for responsibility is an intrinsic feature of agents. There is a tendency in much of the literature on responsible agency to describe its requirements in a way that focuses exclusively on the agent. Responsible agency is thought of as a set of properties that describe an agent in a vacuum, free of a context or environment. We are better off conceiving of the notion of self-governance or self-control as very explicitly tied to facts about agents in contexts. As the Spanish philosopher Ortega y Gasset maintained, I am myself and my circumstances.¹⁴ The basic idea is this: different contexts put pressure on our capacity for self-governance in a variety of ways, and what capacity we have for self governance depends on our environment in a variety of ways. Reminders of this have been one of the clearest upshots of the recent work in exploring the

¹⁴ See José Ortega y Gasset, *Meditations on Quixote*, 1st American ed. (New York: Norton, 1961).

philosophical implications of situationist social psychology.¹⁵ Whatever its implications may be for virtue theory accounts of ethics, it is clear that situational factors can play a large and frequently unappreciated role in our ability to recognize moral considerations and to govern our behavior in light of them. Whether we recently found some small amount of coins, are under time pressure, are smelling cinnamon buns, encountering an object on the right or left, subject to pressure from authority figures, and so on, can play a powerful and often unexpected role in our sensitivity to moral considerations and our capacity for self governance. At a general enough level, this is not news. If I want to avoid stealing music, I would do well to avoid installing software on my computer that facilitates music piracy and increases the temptation to act on piratical impulses. Nevertheless, one of the challenges for an ecologically-sensitive account of moral responsibility—one that acknowledges how our agential capacities work in a range of real world environments—will be to say something about those range of conditions that undermine (and enhance) BASRA, beyond the obvious ones. In some circumstances we will find that we have little or no control over the environmental influences to which we are subject. In other cases we will find that we can exert some control over the environment that influence us, and when we can, this imposes additional burdens on us.

One thing situationist social psychology alerts us to is the implausibility of supposing we have a single, uniform, general capacity to govern our behavior in light of moral reasons, or for that matter, to detect them. I might be strongly resistant to the temptation of excess drinking across a wide range of contexts without being similarly resistant to the temptation to mock my less intelligent students in a wide range of contexts. My capacity to govern myself in light of basic decency considerations might be different than my capacity to govern myself in light of concerns about copyright law, and the force each of these concerns will have on me will presumably vary across

¹⁵ See, for example, John Doris, *Lack of Character* (New York: Cambridge, 2002), Nelkin, “Freedom, Responsibility, and the Challenge of Situationism.”

contexts as well. Thus, rather than conceiving of self-governance as a general capacity, we should think of it as a set of capacities, variably accessible for a given agent to exercise, tied to types of contexts where the agent has the capacity to detect moral considerations (or culpably fails to have that capacity).¹⁶

How hard or difficult it is to exercise a capacity that one has is not an issue directly relevant for the theory of responsible agency. It might be indirectly relevant in that considerations of motivational difficulty might help specify the range or kind of capacity that is relevant for responsible agency. Once we have determined the precise notion of capacity required for BASRA, then the difficulty in exercising that capacity is altogether irrelevant for settling the issue of whether one has BASRA.

This is not to say that motivational impediments (or motivational ease, perhaps) are altogether irrelevant to assessments of moral responsibility. As I will argue below, they are. Where they enter into assessments of moral responsibility is on the side of what the norms of responsibility dictate. Whether the agent is especially blameworthy or praiseworthy for how he or she acts might well be a function of both the moral significance of the act and the difficulty or ease in acting as morality demands. (In this respect, praiseworthiness and blameworthiness can come in degrees.) But the difficulty or ease of what Bernard Williams once called “moral weightlifting” is not directly relevant to the question of whether the agent has BASRA, once we have specified the required capacity for self-governance.¹⁷

¹⁶ This is nothing more than the outline of an argument that needs to be made in greater detail. I intend to provide that argument at a hopefully-not-too-much-later date. However, many of the materials required to make sense of these considerations can be found in the works cited in the previous footnote.

¹⁷ A responsible agent is a self-governed agent, at least in some contexts. I should caution that the notion of self-governance not be construed too broadly, however. There is no reason to expect that the capacity for self-governance required for responsible agency is the same capacity required for self-governance in other contexts. The capacities required for an agent to count as self-governing in political contexts will plausibly be significantly different. And, the sort of self-governance required for “agency at its best” versions of autonomy might be very different than, for example, the kinds of self-governance required for less magnificent forms of agency, including responsible agency. So, we should be cautious of running together the notion of responsible agency with other important forms of agency.

To be a responsible agent is to be the kind of agent that, in the considered context, is sensitive to moral considerations or culpably insensitive, and is capable of governing him or herself, in contexts of that sort, in light of what the norms of responsibility require. As I have argued, it is no easy thing to specify the precise conditions required for having or lacking the capacity for self-governance in light of moral considerations. Nevertheless, conditionalized accounts of ‘can’ present a model of less-than-GFP senses of capacity, of which we can expect a more precise account of the semantics of a prescriptive account of the capacity for agential self-control presupposes. This notion of ‘can’ is indexed to (1) facts about what the aims and demands are of the norms of responsibility, (2) facts about the agent and what considerations he or she has available, and (3) facts about the context and its significance for sensitivity to moral considerations and self-governance in light of those considerations. Together with the sensitivity requirement on moral considerations, the semi-structuralism and BASRA, we have the outlines of an account of responsible agency.

4. What makes this account different from all other accounts?

The account of responsible agency I have sketched shares a great deal with a number of other approaches to responsible agency. What is distinctive about the account I have offered is that it emphasizes (1) the importance of specifically moral considerations, (2) the absence of any single mechanism or faculty involved in the detection of moral considerations in light of (3) the diversity of things that might count as moral considerations. Additionally, (4) the account offers an explicitly revisionist take on what is required to make sense of the notion of a capacity for self-governance, and (5) acknowledges that responsible agency is a thing that is had or not partially in light of the context in which the agent finds him or herself, and (6) relatedly, does not presuppose a single or general capacity for self-governance. These differences give the account a different shape and

different set of dialectical advantages when compared to some prominent alternatives in the literature.¹⁸ Just how it does this is something I will pursue in the later sections of this paper.

One might wonder about the appeal of an account of responsible agency that focuses on moral considerations at all. An important part of the appeal, I believe, connects to what I take to be the most persuasive account of moral responsibility. As I see it, norms of moral responsibility are a special part of the web of moral norms. It is tempting to treat responsibility norms and the underlying social and psychological mechanisms they rely upon as an enforcement mechanism directed at securing our compliance with justified moral norms. At one level, they may well do that. As I mentioned at the start of this paper, it is nevertheless fruitful to conceive of responsibility norms and their related social and psychological aspects as directed at getting us to attend to what moral considerations there are. On this way of conceiving of things, it makes good sense to think of responsible agency as essentially bound up with capacities to be sensitive to moral considerations. Moral evaluation, judgments of praiseworthiness and blameworthiness, morally-loaded reactions such as resentment and indignation, all are sensible and justifiable when directed at creatures who can, at least in principle, track moral considerations, and can govern themselves in light of them. There is a contingent connection between the expression of responsibility-characteristic attitudes at a target, the target's recognition of that expression, and the turning of the target's attention to moral considerations. But this is something that is a commonplace in at least normal adult humans. (Were we to encounter creatures unmoved by expressions of these attitudes, but that were nonetheless capable of recognizing moral considerations and governing their behavior in a way that tracked those considerations, these agents would be responsible agents, but not agents at whom expressions of responsibility-characteristic attitudes would be fully sensible.) At any rate, creatures that altogether

¹⁸ The account I offer differs from Fischer and Ravizza's account on at least (1), (2), (4), and (I think), (5). The account differs from Wallace's account on at least (4), (5), and (at least as I read him) (6) and perhaps (2). To my knowledge, few accounts have spoken one way or another on (3).

lack the sensitivity to moral considerations are poor targets of responsibility-characteristic evaluation, in part because these practices lose their point, from both a pragmatic and a moral standpoint. Moral exhortation directed at a beast that can neither track moral considerations nor govern itself in light of them is a waste of time, and (at least typically) unjustified from a moral standpoint.¹⁹ Construing responsible agency in terms of capacities centered on moral considerations is therefore partly a function of the aim of a system of responsibility.

A second reason for favoring this sort of account is much simpler: if you accept the idea that moral considerations and self-governance in light of them are compatible with determinism²⁰, then this is an appealing way of making sense of how we *ought* to think about responsible agency, without invoking what P.F. Strawson memorably called “panicky metaphysics.” My aim has been to propose the core of an account that can do the work we expect of a theory of responsible agency at a minimum of ontological cost.

With this framework for thinking about responsible agency in place, I now want to consider some alternative pictures. In particular, I will focus on whether my account can capture some of their insights while avoiding some of the difficulties they raise. I will focus on three issues that are prominent in alternative approaches to responsible agency: the attractions of an identificationist

¹⁹ I say “typically” to acknowledge two wrinkles. First, one way to get at least some creatures to track moral considerations is via a process of habituation. We typically call these creatures ‘children’ and the habituating process ‘doing a good job of childrearing’. In cases where the development of this sort of moral sensitivity isn’t possible, then it is much harder to see what the justification might be for distinctly moral (and especially responsibility-assessing) exhortation might be, although we might find that some non-moral norms gain efficacy when they co-opt moral attitudes. The second wrinkle concerns creatures whose inability to track moral considerations and to govern themselves in light of them might be limited to some contexts and not others. Just what this amounts to will have to await specification of the relevant capacities, but we should not rule out the possibility that the relevant notion of capacity could turn out to be highly context-specific.

²⁰ For example, even if determinism is true, you might help me because you see that I need help. In doing so, you would be (among other things) responding to the consideration generated by my being in need. These things might be analyzable in terms of more basic things, but it does not appear to affect the truth of the claim “you helped me because I needed help.” There is wide consensus about this idea, at least with respect to moral reasons. See, for example, Hilary Bok, *Freedom and Responsibility* (Princeton, N.J.: Princeton University Press, 1998), Daniel Dennett, *Elbow Room* (Cambridge: MIT, 1984), Daniel Dennett, *Freedom Evolves* (New York: Viking, 2003), John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control*. (Cambridge, Mass.: Blackwell, 1994), Derk Pereboom, *Living without Free Will* (Cambridge: Cambridge, 2001).

account of responsible agency; the importance of a developmental model of responsible agency; and the importance of (libertarian) efforts of will.

5. Identification, Tracing, and Manipulation

In this section I want to consider an alternative picture of responsible agency, one that emphasizes that an agent is morally responsible when the agent identifies with the values or motives that lead to the action.²¹ On these views, an agent's BASRA is, roughly, determined by the possession of the capacity to identify with some of one's own values or desires.

Recently, Doris and Stich have argued that common sense is not really incompatibilist, as one might have thought on the basis of thought experiments alone. The Doris and Stich view is motivated by a series of experiments conducted by Woolfolk, Doris, and Darley. What these experiments seem to show is that attributions of responsibility tend to track an agent's identification with the action; identification or its absence is, under normal circumstances, the most salient trigger of our assessments of responsibility.²²

If Doris and Stich are correct, this looks problematic for my account, as I have not focused on identification as the marker of responsible agency. If Doris and Stich are right my account is either gratuitous, mistaken, or both. It may even be the case that if we simply understand the concepts we do have, there is no need to accept revisionism at all.

Here's a synopsis of the case crucial to the experiments discussed by Doris and Stich: There are two couples that are friends, returning from a vacation together. Bill has learned that his wife (Susan) and his best friend (Frank) have been involved in a love affair, of which Bill has just

²¹ This sort of view is most famously associated with the work of Harry Frankfurt, but aspects of it have received development in a number of places, including Michael E. Bratman, "Identification, Decision, and Treating as a Reason," *Philosophical Topics* 24, no. 2 (1996), Gary Watson, "Free Agency," *Journal of Philosophy* 72, no. 8 (1975).

²² Doris and Stich, "As a Matter of Fact: Empirical Perspectives on Ethics," Robert L. Woolfolk, John Doris, and John Darley, "Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility," *Cognition* (forthcoming). David Hume's treatment of responsibility is perhaps the most recognized identificationist account. See David Hume, *A Treatise of Human Nature*, 2 ed. (New York: Oxford University Press, 1978).

discovered proof. In the low-identification version of the case Bill decides that he is going to confront Susan and Frank, but that he has resolved not to stand in their way if they want to be together. In the high-identification version of the case, Bill decides that he will kill Frank. Here's where things get philosophically interesting. Before Bill does anything, hijackers take over the plane and things eventually get to a situation where Bill is ordered by the hijackers to shoot Frank, and he does so. What Doris and Stich report is that "the high identification actor was judged more responsible, more appropriately blamed, and more properly subject to guilt than the low identification actor" (§5). That is, when Bill identified with killing Frank, subjects hold him more responsible and more frequently hold him responsible than they did in scenarios where Bill did not identify with killing Frank. Strikingly, this was the case even when the hijackers administered to Bill a "compliance drug" that forced him to behave exactly as they ordered. If Bill strongly identified with killing Frank, it does not seem to matter whether or not he could have done otherwise at the time he pulls the trigger.²³

While these results are provocative, it is not clear that this evidence should worry incompatibilists who claim that incompatibilism is deeply entrenched in common sense, and, I think, it does not provide a compelling reason to adopt an identificationist account of responsible agency. (For that matter, it should not worry compatibilists who think that moral responsibility requires alternative possibilities, construed in some compatibilist-friendly way.)

A standard aspect of many contemporary theories of free will is reliance on a principle of tracing. The idea of tracing is that some outcome for which one is responsible need not be anchored in an agent or an agent's action immediately prior to the outcome, but rather some suitable time prior to the moment of deliberation or action. For example, suppose I have a regular policy of insulting any student who comes to office hours. Suppose that I am no longer reflective at all about

²³ This scenario was tested precisely because of concerns that in less coercive versions of the case, there remained alternative possibilities that might fuel an incompatibilist reading of the evidence.

my practice, and not sensitive to moral considerations that weigh against it. When a student arrives, I just say “What stupid question are you too dumb to answer on your own?” On any theory that involves tracing—compatibilist or incompatibilist—the most natural thing to say about this case is that what makes me responsible for insulting my students is not that I was paradigmatically free in my action at the relevant times. How could I be? I was simply acting out a script I had long since settled on and followed automatically. More dramatic cases involve drunk driving. A drunk driver does not get off the moral hook simply because at the time he hit someone with his care he was especially intoxicated. In these kinds of cases, it makes more sense to think of my responsibility for the insult as grounded in some early decision or combination of decisions to, for example, insult students in this fashion or to start drinking when I knew I would be tempted to drive.

Incorporation of an idea of tracing has been a staple of a range of theories, and once we recognize it as a possibility it becomes harder to see how examples of the sort cited by Doris and Stich will be problematic for incompatibilists (or really, anyone not an identificationist) to handle. Non-identificationists will simply insist that Bill’s responsibility for his killing Frank is grounded in his freely acquired decision to kill Frank, prior to the actions of the hijackers. While Bill might not have envisioned the particular details of how we was going to kill Frank, his deciding to do so might be a sufficient anchor for an incompatibilist trace. As long as there is no reason to suppose the prior decision violated incompatibilist conditions of responsible agency (and nothing in the prompt suggests otherwise), then there is no reason to rule out this sort of tracing. That the hijackers coerced Bill might involve some diminution of responsibility—this is consistent with the responses Woolfolk, et.al. received—does not mean that Bill cannot be held morally responsible for pulling the trigger. This is not to deny that identification may play some appropriate roll in our attributions of responsibility. It might, for example, tell us something about the extent to which an agent manifest ill will, or whether an agent is showing the appropriate degree of moral concern for his or her

actions. Identification might also be important mechanism for getting agents to take responsibility, and to see themselves as appropriately connected to the action in a way that justifies and merits responsibility. What these considerations and the empirical evidence fail to speak to is whether or not there is some further incompatibilist condition lurking in ordinary attributions of responsibility.²⁴

A second reason for caution about the evidence that Doris and Stich rely on concerns another feature of responsibility attributions. In a different set of experiments, Nichols and Knobe discovered that responsibility attributions are sensitive to the way a case is framed. In concrete, high affect contexts, attributions skew compatibilist. However, when a case is discussed abstractly, in low affect terms, responsibility attributions strongly track incompatibilist (particularly libertarian) conceptions of agency.²⁵ What this evidence suggests is that cases like those described in the Woolfolk et. al. experiments will tend to yield compatibilist answers, even in high constraint contexts. Perhaps an identificationist account is the correct account of how responsibility attributions work in certain contexts. What the Nichols and Knobe evidence suggests is that this is not the complete story, for there are contexts in which our attributions seem to focus on more

²⁴ Doris and Stich seem to think that the evidence also speaks against forms of incompatibilism that do not require alternative possibilities—what Michael McKenna has helpfully dubbed “source incompatibilism”—I do not see how this can be so. According to source incompatibilists, the removal of alternative possibilities does not, by itself, mean that the agent wasn’t the ultimate source of the action. It is difficult to see how one might be an ultimate source without alternative possibilities, but this is precisely the lesson that some source incompatibilists have tried to draw from Frankfurt-style counterexamples to the Principle of Alternative Possibilities (see, for example, Derk Pereboom, “Defending Hard Incompatibilism,” *Midwest Studies in Philosophy* 29, no. 1 (2005).). Conceivably, a source incompatibilist might argue that Bill *was* ultimately responsible (assuming he wasn’t subject to causal determinism), and that his ultimate responsibility was not gotten rid of simply because he lacked alternative possibilities. We would need some account of what Bill’s sourcehood consists in, but I don’t see any obvious reason why the case of Bill *prevents* source incompatibilists from offering an account compatible with the case as it has been described. (Compare the case they rely on with one where the hijackers give Bill a pill that deterministically makes him identify with whatever action they give him. If Bill didn’t previously identify with the action, this sort of coercion would clearly undermine source-hood. I wager it would also undermine the rate at which respondents attribute moral responsibility. Again, though, perhaps a source incompatibilist could find room for libertarian agency even in this picture.) Whatever the success or failure of this line of source incompatibilist response might be, it should be clear that there is no reason a source incompatibilist could not help him or herself to a tracing approach, and thus dodge the consequences of the Woolfolk, et. al., evidence in this way.

²⁵ Nichols and Knobe, “Moral Responsibility and Determinism: Empirical Investigations of Folk Concepts.”

traditional incompatibilists concerns about alternative possibilities, understood in a way that hold fixed the past and the laws of nature.²⁶

It will strike many as good news that this new argument for identificationism has a ready reply. Identificationist views have raised worries about their difficulty in handling the possibility of certain forms of manipulation. In particular, identificationist views appear to be committed to the possibility that someone can be responsible, even when that person has been brainwashed into identifying with particular actions, desires, or values. For many philosophers, this is unacceptably counterintuitive.

An identificationist could try to bite the bullet, accepting that (at least in principle) we could be both responsible and manipulated or brainwashed into the state that makes us responsible. While I think there may be some not-entirely-obvious attractions to the view, it is *prima facie* implausible when identificationism is presented as an account of our folk concepts. Worries about systematic and profound manipulation are too widespread and deep to be plausibly characterized as a misconstrual of our understanding about the conditions for excusing or exemption from responsibility. At any rate, it would take a substantial body of perhaps experimental work to convince many philosophers that we do not ordinarily think that manipulation or brainwashing of a deep and systematic sort is not problematic for responsible agency. Perhaps it *ought* not pose serious problems, but that would be a different argument than the one suggested by Doris and Stich's discussion.

Identificationist conditions may reflect something important about how we come to know whether or not someone is a responsible agent, and they may play some role in our actual ascriptions

²⁶ This is a possibility highlighted by Woolfolk et. al., but not considered in Doris and Stich. For my part, I am inclined to think that we probably can elicit some genuine compatibilist intuitions in experimental contexts. However, I think that (1) any such intuitions typically reflect pragmatic or epistemic aspects of our practice of responsibility ascriptions and not anything about the folk metaphysics of responsibility, and (2) our ongoing ability to elicit incompatibilist intuitions counts as a victory for incompatibilist accounts about folk concepts of freedom and moral responsibility. I say more about these issues in Vargas, "Philosophy and the Folk."

of responsibility. Indeed, we could conceivably grant that identification in some sense is a necessary condition for responsibility; at any rate, it does not appear to undermine responsibility when the elements of BASRA are in place. It is a mistake, however, to treat identification as the central feature of the metaphysics of responsible agency, where its presence or absence, along with ordinary agential capacities, is sufficient for responsibility. In reply to manipulation worries, I am prepared to acknowledge that our current folk concepts of responsibility and responsible agency may well rule out the possibility that one can be manipulated or brain-washed into responsible agency. Elsewhere, I have tried to show that once we get clear on the aim of a system of moral responsibility (i.e., fostering agents who are sensitive to moral considerations) and once we acknowledge that the chief issue is to correctly characterize the view we *ought* to have of moral responsibility, and not merely the concept we have inherited, we should adopt what I have called a semi-structural view of responsible agency. On the semi-structural view, our concern is whether an agent has the basic structure of responsible agency (including the capacity to recognize and respond to moral considerations) in the relevant context, and if not, whether the agent was responsible for its absence. On this sort of view, it might well be possible to brainwash or manipulate someone into responsibility. The important thing to focus on is that not just any manipulation is compatible with responsible agency. To manipulate or brainwash someone into responsible agency would require the introduction of very special capacities—the very capacities that make it appropriate to evaluate the agent in a moral light. None of this is to deny that we may, as a matter of the moral concepts we find ourselves with, think that any sufficiently deep or systematic piece of manipulation renders us non-responsible. My point has instead been that if we have the right understanding of responsibility and the role that responsible agency plays with respect to the norms of responsibility, we ought not be so concerned about manipulation that turns us into responsible agents. It is, in general, what we might hope others and the world will cooperate in doing for us.

6. Development and tracing troubles

I now want to return to the issue of tracing, and its relationship to a *developmental model* of responsible agency. A developmental model is one where we gradually become morally responsible for an increasing number of things in a wider range of contexts. There is substantial appeal to this model. It is motivated by the idea that no one starts a life being fully responsible, like the Athena of mythology, a fully responsible agent *ex nihilo*.

The insights that give rise to a developmental model dovetail with the considerations that favor adoption of a tracing theory. One way in which we are responsible for what we do is by being responsible for who we are. A good number of our actions issue from traits, habits, and policies that are partly constitutive of who we are. It is by being responsible for these things that we come to be responsible for much of what we do, and this happens over extended periods of time. Part of what the idea of tracing does, then, is to give us a way of making sense of this idea. Our responsibility for our selves is something that builds up, in pieces, over time. The incorporation of tracing in a theory of responsibility reflects the logic of our thinking about how we come to be responsible for things whose most immediate causal source is our habits, traits of character, or standing policies about what to do.

Although both developmental models and the role of tracing in them are common in the literature of responsible agency²⁷, these accounts are plagued by an under-appreciated difficulty. The difficulty is this: in many circumstances, traits, habits, or policies are adopted under conditions in which we have poor epistemic access to the later consequences that will follow from having adopted

²⁷ See, among others, Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, Robert Kane, *The Significance of Free Will* (Oxford: Oxford, 1996). Many philosophers rely on tracing. See Laura Waddell Ekstrom, *Free Will: A Philosophical Study* (Boulder, Colorado: Westview Press, 2000), Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, Kane, *The Significance of Free Will*, Peter Van Inwagen, "When Is the Will Free?" in *Philosophical Perspectives*, 3, *Philosophy of Mind and Action Theory*, 1989, ed. James E. Tomberlin (Atascadero, CA: Ridgeview, 1989).

that trait, habit, or policy. Moral responsibility for some outcome, however, seems to require something like knowledge about that outcome being foreseeable. It is almost a truism that I cannot be held responsible for some outcome unless it was reasonably foreseeable—except where my lack of foresight is itself something for which I am responsible. In the context of developmental models, the worry is this: in a wide range of cases, the aspects of our self, character, or policy which provide the basis for many of our actions were acquired under circumstances where we could not foresee our acting in the appropriate way. Or, to put it somewhat differently, the anchors for our responsibility traces cannot secure responsibility when the downstream effect was not reasonably foreseeable at the time of the anchoring decision. Indeed, the more remote—temporally or recognitionally—the context of action is from the context of the acquisition of the trait, habit, or policy, the more significant we should expect the epistemic defect to become. Many of the characteristics I inculcated in myself in high school, for example, and have since forgotten about were doubtlessly acquired under conditions when I would or could not know about their consequences in more mature adult life.²⁸

How ubiquitous this problem is remains an open question. As a problem for theories of responsibility, it depends in part on the frequency with which the theory relies on tracing. *Restrictivist* accounts that hold that we have free will somewhat infrequently will face a version of this problem to a greater extent than theories that require little or no tracing, or whose tracing does not typically involve significant temporal extendedness.

For accounts that emphasize an agent's sensitivity to moral considerations, the problem strikes me as less pressing. Even habitual actions can be sensitive to moral considerations. That I habitually empty my pockets on a bookshelf when I get home from work does not mean that, were there something I perceived as more important, that I would not respond to those considerations.

²⁸ This paragraph sketches an argument I have developed in more detail in Manuel Vargas, "The Trouble with Tracing," *Midwest Studies in Philosophy* 29, no. 1 (2005).

To be sure, our habits, traits, or policies might in some circumstances make us less able to detect some moral considerations. In other cases, though, they improve it. If I had no habit of asking my children how their day went, I would presumably fail at least some of the time (and depending on facts about me and my children, perhaps most of the time) to be aware of some considerations that should weigh in my deliberations.

Tracing might prove to be a more systematic problem for this sort of account if one thought that responsible agency was historical in some deep and systematic sense. Fischer and Ravizza's account of reasons-responsiveness has this feature. They argue that irrespective of what one's reasons-responsive capacities might be, there will always be some historical condition that must be satisfied for one to be a responsible agent.²⁹ On accounts such as these the historical ownership condition will introduce an element that, at least in principle, seems susceptible to the sort of difficulty I have outlined with tracing pictures.

These considerations leave an account of responsible agency with a challenge. How are we to make sense of the idea that our being responsible agents is a result of a temporally-extended, personal developmental history without invoking the responsibility-undermining trouble that tracing seems to bring with it? The traces get a significant degree of temporal extendedness in any developmentalist story that emphasizes agents having particular histories, and this is where the problem seems most profound. The solution is not to give up on the insight that our being responsible agents is something that does not happen over night. The mistake seems to be in accommodating this insight by appeal to early decisions in an agent's life doing a significant part of the anchoring of for downstream traits, habits, decisions, and so on. Instead, we should hold that our basic agential structure is context-dependent and semi-structuralist.

²⁹ This is, of course, another important way in which my account differs from theirs, beyond my emphasis on moral considerations (in contrast to their emphasis on reasons— moral or otherwise), my commitment to revisionism, and my focus on context-sensitivity.

The way to respect the fact that we get our responsible agency gradually is to acknowledge that we are not universally good at recognizing moral considerations, and that our ability to appropriately act on moral considerations is partly a function of the environment we find ourselves in. Moral development is largely about increasing the range of contexts in which we do have due moral sensitivity, and widening the range of contexts in which we can appropriately respond to our sensitivities. This is how and why we can think of our children as responsible for some actions and not others. Even if we would not think of ordinary children as being responsible in the majority of contexts faced by ordinary adults in commonplace contexts, we can think that there are limited contexts (initially none, later more) in which a child can come to be responsible. Of course, a great deal of our moral interactions with children are directed at moral education. There are plenty of instances in which we don't so much blame as much as we emulate blame in an attempt to get our children to start tracking those considerations worth attending to. Still, there are contexts in which we genuinely blame and genuinely praise. The mistake would be in supposing that this signals something like a stable, cross-situational capacity to recognize moral considerations and to guide conduct in light of those considerations.

The semi-structuralism buys the account a limited exposure to the need for tracing. The principle cases in which tracing will be invoked is in cases in which one lacks BASRA. It seems plausible to think that the tracing trouble having to do with reasonable foreseeability does not frequently occur in cases where one has intentionally brought it about that one destroyed one's basic structure of responsible agency.

The revisionism of the account also buys something: the account has some degree of flexibility with respect to counterintuitiveness. To the extent that the account seems counterintuitive

because it is not deeply historical in its requirements for responsible agency, we need not worry.³⁰ That said, there is already some reason to think that we do not think *all* manipulation undermines responsibility. As I noted before, good parenting sometimes amounts to the manipulation and introduction of responsibility-supporting capacities in an agent that lacks them or has them in underdeveloped forms.

One lesson from all of this is that there are two families of considerations that independently fund the view that responsible agency should be thought of as semi-structural. Considerations about history and the aim of responsibility suggest that a prescriptive account of responsible agency should be semi-structural, and considerations about the trouble with tracing suggests the same. The good news is that we can do this without giving up on a developmental model of moral responsibility. Although some amount of tracing will be necessary when BASRA is absent, a context-sensitive semi-structural account that emphasizes the grasping of moral considerations and their appropriate response can limit how much tracing must be invoked.

7. Libertarian efforts of the will

One could accept most of the account of responsible agency I have offered and grant that it need not worry about identificationist and tracing threats, but one might still worry that the account I have offered fails to accommodate an important aspect of moral responsibility: the role for efforts of will, and in particular, libertarian efforts of will. In this section, I consider a contrasting account that does emphasize the role of libertarian efforts of will—Robert Kane’s justly influential account—and I argue that its difficulties and successes are instructive for the account I favor.

³⁰ Of course, the account is not immune to *any* sort of counterexample or any instance of counterintuitiveness. It is still governed by constraints of coherence, naturalistic plausibility, and normative adequacy. Some account would also be required in any instance where the counterintuitiveness could not be traced back to some aspect of our folk conception of responsible agency that is not being abandoned for good reason. In this case, I have argued that the good reason for giving up a systematically historical requirement is that it does no normative work (that is, we are requiring something for no morally or ethically relevant reason).

On Kane's account, responsible agency requires indeterminism in "self-forming actions" (SFAs), which are actions or willings that are indeterministic, intentional, voluntary, and endorsed by the agent in circumstances where he or she faces a choice between competing motivations. He describes these moments in the following way:

There is a tension and uncertainty in our minds at such times of inner conflict that is reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium—in short, a kind of stirring up of chaos in the brain that makes it sensitive to micro-indeterminacies at the neuronal level. As a result, the uncertainty and inner tension we feel at such soul-searching moments of self-formation is reflected in the indeterminacy of our neural processes themselves. What is experienced phenomenologically as uncertainty corresponds physically to the opening of a window of opportunity that temporarily screens off complete determination by the past

. . . . When we decide in such circumstances, and the indeterminate efforts we are making become determinate choices, we *make* one set of competing reasons or motives prevail over the others then and there *by deciding*.³¹

Kane supposes that there are at least two recurrent, connected, and competing neural networks in an agent, with each network containing as its input a desire, motivation, or consideration and as its output some choice that satisfies the desire. The networks are competing because the agent cannot choose to satisfy both desires, and the satisfaction of one precludes the satisfaction of the other. Since the networks are connected and chaotic, according to Kane, the conflict between them makes them susceptible to lower-level indeterminacies. When the agent decides on one option over the other, this corresponds to one of the neural pathways reaching an

³¹ This passage is taken from pp. 224-225 of Robert Kane, "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism," *Journal of Philosophy* XCVI, no. 5 (1999). See also Robert Kane, "On Free Will, Responsibility and Indeterminism: Responses to Clarke, Haji, and Mele.," *Philosophical Explorations* 2 (1999).

activation threshold, overcoming the indeterminism of the other.³² And of course, for this to be a picture of responsible agency, all the standard conditions apply— e.g., the relevant choices are not implanted by “nefarious neurosurgeons,” the agent will not fail to endorse the outcome, disassociating him or herself from the choice, and the resolution of choice is voluntary and intentional.

Agents who have SFAs of the sort Kane describes are not doing things by mistake, accidentally, inadvertently, or involuntarily. Whatever choice the agent makes, is voluntary, intentional, endorsed, and follows from the motivations of the agent. While it may be true that an agent and a duplicate with the same histories might diverge after *t*, the way in which this divergence is brought about is not, he maintains, responsibility undermining as some might think. Kane makes several points in defense of this claim:

1. The presence of plural and competing tasks in the brain does not entail two people or a disassociated self.
2. That there is indeterminism in the deliberation does not make the deliberation “unowned” by the agent. The indeterminism comes from the agent’s own will, from the agent’s desires for two (or more) mutually exclusive outcomes, and by stipulation, either outcome would be endorsed by the agent.
3. The indeterminism cannot be separated from the effort to decide between options— the indeterminism is a property of the effort to decide between options.
4. Whichever choice results, the agent was “trying to make both choices.”³³
5. The agent can be held responsible for whichever of the choices result because the agent will have succeeded in doing what he or she was trying to do.³⁴

³² Ibid, p. 226.

³³ Kane, “Responsibility, Luck, and Chance” p. 231.

Kane adopted the “parallel processing” picture in response to what has become known as the *Luck Argument*. In its most basic formulation, the Luck Argument holds that for any undetermined action or willing, whether it happens or not is a matter of luck, and therefore not an instance of free and responsible action. Libertarianism about free will requires indeterminism, but (as the argument goes) since indeterminism amounts to luck, and luck undermines responsibility and freedom, libertarianism is responsibility undermining.³⁵

While I am agnostic about whether the Luck Argument is really a problem for Kane and other libertarians³⁶, I do have independent doubts about the picture of responsible agency he offers in response to it. Here is what I take to be the difficulty: The power of Kane’s response to the Luck Argument seems to hinge on a misdescription of the situation. Consider Kane’s analysis of the agency he describes: “if she nevertheless *succeeds*, then she can be held responsible because, like them, she will have succeeded in doing *what she was trying to do*. And the interesting thing is that this will be true of her, *whichever choice is made*, because she was trying to make both choices and one is going to

³⁴ Kane, “Responsibility, Luck, and Chance” p. 233.

³⁵ Although its details are not strictly necessary for present purposes, the version of the Luck Argument Kane is primarily considered with is essentially the following, where P= an agent, A= an action, and t= a particular time:

1. In the actual world, P voluntarily and intentionally does A at t, where A is undetermined at t.
2. In a possible world, which is the same as the actual world up to T, P* (P’s counterpart with the same past) voluntarily and intentionally does otherwise (does B) at t.
3. But then (since their pasts are the same), there is nothing about the agents’ powers, capacities, states of mind, characters, dispositions, motives, etc. prior to t that explains the difference in actions, including choices, in the two possible worlds.
4. It is therefore a matter of luck or chance that P voluntarily and intentionally does A at t, whereas P* voluntarily and intentionally does B at t.
5. P is therefore not responsible for voluntarily and intentionally doing A at t (and presumably P* is also not responsible for voluntarily and intentionally doing B at t).

This version of the argument is modeled after the one offered in Alfred Mele, “Kane, Luck, and the Significance of Free Will,” *Philosophical Explorations* 2 (1999), Alfred Mele, “Review of the Significance of Free Will,” *Journal of Philosophy* XCV, no. 11 (1998), Alfred Mele, “Ultimate Responsibility and Dumb Luck,” *Social Philosophy and Policy* 16 (1999). For Kane’s presentation of Mele’s argument, see “Responsibility, Luck, and Chance,” pp. 229-230. Other versions of the argument are also cited in footnote 10 of “Responsibility, Luck, and Chance.”

³⁶ There is an interesting literature on whether the Luck Argument is really a problem for libertarians. See, for example, Randolph Clarke, “Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency,” in *The Oxford Handbook of Free Will*, ed. Robert Kane (New York: Oxford, 2002). See also L. Ekstrom, “Free Will, Chance, and Mystery,” *Philosophical Studies* 113, no. 2 (2003).

succeed”³⁷ (this is claim 4, above). Kane’s description of the agent as trying to make both choices seems to be a misdescription. What the agent is doing is trying to make *a* choice. We might concede that the agent had desired to do two different things, and that the agent may have two competing sub-agential processes at work in his brain (the competing neural networks), or even that the agent could be colloquially described as “being of two minds.” However, these facts do not count towards the agent trying to make both choices. Rather, what the agent is trying to make is *a* choice, and the competing desires (or their neurological correlates) were elements that played a role in that decision. But even if we say of these sub-agential elements that they were “trying” to do different things, it does not follow that the agent was trying to do both of those things. Even if having competing desires (or having competing neural networks) entails that sub-agential processes are “trying” to do different things, it is surely false that the *agent* was trying to do both things.

Here is why: suppose that an agent has five (why stop at two?) neural networks in operation that have as their outputs at a given decision point the following outcomes: jumping off a cliff, shooting a breathtaking photo, making a marriage proposal, resolving to pursue the life of a profligate debauchee, and shouting his favorite cartoon catchphrases at the top of his lungs. On Kane’s view, these are the choices the agent is trying to make. Let us suppose the agent now decides to propose marriage. If Kane’s account is to be believed, the agent may have decided to propose, but it was also true that he was trying to make the other choices. But while the agent may have wanted (in addition to marriage) to jump off a cliff, resolve to become a profligate debauchee, shout his favorite cartoon catchphrases, or take a picture, he surely did not try to do any of these things. The only thing he tried to do was to make a choice. Given the choice he did make, it thereby became true to say that he tried to get married. However, the truth of his tryings is determined by his decisions, not by his wants (or if you like, the actions outcomes of his neural networks). Agents

³⁷ Kane “Responsibility, Luck, and Chance,” p. 231 [italics in the original].

almost never try to make every choice they would like to make. So, Kane cannot be right when he claims that the paradigmatically responsible libertarian agent is trying to make both choices.

The argument that it is misleading to describe the activities of the interacting neural processes as “tryings” might provoke an attempt at an easy amendment. We might amend Kane’s proposal by specifying that ‘tryings’, etc., are terms of art, made to pick out processes that stand in the appropriate relation to the agent. But if we do this, then it is not clear that Kane’s argument maintains its persuasiveness, for the original point was that the decision was somehow settled by the agent’s trying, understood in a robust sense. Changing that sense of “trying” merely papers over the underlying problem of attributing agent-level powers to sub-agential processes (whether they be the operations of neural networks, biochemistry, or quantum physics).

One could be tempted to simply change the problematic claim to read:

(4*) Whichever choice results, the agent owned the process that led to the choice.

Though 4* is not obviously vulnerable to charges of a category mistake or misdescription of the phenomenon, it is not clear that it (with the other elements of his account) entitles Kane to concluding that the agent’s choice was not a matter of luck. Given that on Kane’s account, one condition of ownership is that the agent at some point had to have an SFA that led to its becoming the agent’s own process, then the worry about luck can just regress back through the sequence of SFAs that lead to ownership. At the end of the regress, it looks like some degree of luck is once again waiting, or failing that, the tracing trouble I discussed in the previous section.

I am inclined to think that we should acknowledge some degree of luck in the constitution of our agency, but go on to hold that this issue does not speak to whether or not an agent has BASRA, and thus, whether the agent is appropriately subject to the norms of moral responsibility. It

might be a matter of luck whether or not one is a responsible agent, but this fact should not be viewed, at least not by itself, as responsibility undermining.

Let us, however, consider a more basic question: what is the motivation for Kane's adoption of the parallel processing picture, as opposed to some other account of libertarian agency that promises to defeat the Luck Argument? The answer is suggested by his earlier view, which relied on the idea of single indeterministic process. It appears that all along his aim was to characterize indeterministic agent-level efforts. In the case where there is a single desire the agent is acting on, it makes sense to identify the agent's efforts with the single process.

(As we have seen, when Kane doubles the desires it does not follow that this doubles the agent's efforts—it merely means there are more things the agent desires to do. The agent is still only trying to do one thing: make a decision about what to do.)

Although the move to sub-agential processes does not get Kane what he wants, there is something important about trying to locate indeterminacy in agent-level phenomena. Many—though certainly not all—philosophers have thought that the indeterminacy required by libertarianism must somehow be linked to normatively salient features of the *agent*. Without some connection between the postulated indeterminacy and the normatively salient features of the agent, any such libertarian proposal will seem particularly vulnerable to the Luck Argument, or failing that, the complaint that the indeterminism is gratuitous.³⁸ Therefore, tying the indeterminacy to the agent's efforts seems like the right place to do it, if you are going to require indeterminism, because an agent's efforts seem to tell us something about the moral worth of the agent.

I am inclined to think we can get along without indeterminism, although it does cost us something by way of our self-image. Still, the idea that an agent's efforts tell us something about the

³⁸ Cf. Gary Watson "Free Action, Free Will" *Mind* 96 (1987), 145-72. For a probing discussion of these issues, especially with an eye towards the issue of control, see Clarke, "Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency." Concern for normatively salient agent-level features presumably plays some role in the appeal of agent causation accounts of libertarianism.

moral worth of the agent is an insight we should hold on to. If we can find a way to make sense of how this might work, it offers something of a reply to the worry that what we do is merely another causal upshot of the universe. If the universe has to go through us, if our efforts of will come into play, and if they matter in a special way for some agents and not others, then this is an important difference between us, non-moral agents, and the non-agential parts of the universe. It might not be the difference we normally suppose, but it is enough to ground evaluations of responsibility.³⁹

On the account I have offered, we can make perfectly good sense of how this works; the freedom condition on moral responsibility will be located in the efforts of will. Recall that on my account, an agent is not responsible unless he or she has, among other things, the capacity to recognize moral considerations and to govern his or her conduct in light of them. This is what distinguishes morally responsible agents from non-moral agents. Where efforts of will come into play is with an agent's endeavoring to appropriately guide his or her conduct in light of those moral considerations. When, as a responsible agent, I fail to guide my conduct in a way sensitive to what moral considerations there are, I become open to moral criticism; I can appropriately be blamed. When I succeed in governing my behavior in light of moral considerations, I am appropriately praised. The structure of our agency and the efforts of our will are what jointly make our responsible actions more than *mere* causal upshots of antecedent forces. Our contribution is distinct and delicate, the kind of thing that both situations and persons can destroy. It is, nonetheless, a real contribution unique to us. We are, in this respect, like works of art, built up out of more basic things, but with properties—beauty, composition, and so on—that are only appropriately described at the level of the work of art. It is in these unique aspects of a painting, *as a painting*, that its contributions to the world are properly understood. Similarly, it is the distinctly agential aspects of us in which our

³⁹ My thoughts on these issues have been influenced by the discussion in John Martin Fischer, "Free Will and Moral Responsibility," in *Oxford Handbook on Ethics*, ed. David Copp (Oxford: Oxford, forthcoming). I suspect that I am somewhat more inclined than Fischer to think that the importance of our indispensability and its suitability for grounding moral evaluation may not be what we were looking for, but it is all we are likely entitled to.

unique contribution to the world is to be understood. For issues of moral responsibility our unique contribution is the effort of willing.

We need not be indeterminists to acknowledge that the freedom condition on responsible agency is appropriately located in the efforts of the will. Indeed, recognizing that this important aspect of our self-image and its connection to moral evaluation is, in principle, separable from a libertarian conception may reduce some pressure to require indeterminism in a revisionist account of responsible agency. To be sure, the sense of freedom or capacity invoked will not be our pre-philosophical vision of GFP-freedom. As we have seen, however, there are many ways to reconstruct the responsibility-relevant sense of freedom, some of which are compatibilists senses of capacity or freedom that we already enjoy and rely upon.⁴⁰

So, what difficulty there is in making sense of the idea that freedom is tied to our efforts of will need not be abandoned because of the problems it raises in a libertarian context. Instead, this insight should be preserved and incorporated in any plausible theory of responsible agency. Whether we need libertarian agency for some further reason is a distinct issue which I will not try to address here.

8. The Better Beast

If my account is essentially correct, moral responsibility is at base about building better beasts. As a system of norms and coordinating practices, it is concerned to foster in us greater sensitivity to moral considerations and better responsiveness to them in a wider range of contexts. What a theory of responsible agency does is to identify which beasts are worth trying to build better.

⁴⁰ In the face of these sorts of considerations, and worries about the general plausibility of libertarianism, some libertarians will suggest that skepticism about free will is the right sort of move. I'm skeptical that for any account, if you think it is the account that best captures commonsense, that a demonstration that we aren't agents of that sort shows that we aren't responsible agents. It might. But I don't think anyone has shown it, and anyway, I think there is good reason to be skeptical about *these* kinds of arguments. See Manuel Vargas, "Libertarianism and Skepticism About Free Will: Some Argument against Both," *Philosophical Topics* 32, no. 1-2 (2004 (forthcoming)).

In this paper, I have endeavored to say something about the kinds of beasts we should understand our system of responsibility—the norms, practices, and characteristics attitudes—as being directed at building. I have endeavored to offer an account of responsible agency that fits with a comprehensive view about the aim of moral responsibility and what the significance of responsibility might be within a larger normative framework. By emphasizing our capacity to be sensitive to moral considerations and to govern our conduct in light of them, we can make sense of how responsibility is related to other moral concerns and why it is that responsible agency seems as important as it does. I have also endeavored to show how the view I favor accommodates a number of traditional concerns for a theory of responsibility (e.g., finding a place for efforts of the will, and for the idea that our achievement of responsible agency occurs over time), and how it addresses a variety of challenges to contemporary theories of responsible agency (e.g., history, manipulation, troubles with tracing, etc.). I have also highlighted what I take to be distinctive about this account and how it captures insights that motivate alternative approaches.

Once we accept a (likely) revisionist, moral consideration-sensitive, semi-structuralist account of responsible agency, we are better suited to appreciate how it can make sense to draw a distinction between what we do and what it is that the universe does. Our freedom may not be the freedom we pre-philosophically suppose it to be. This does not mean that our freedom and our agency are not real or not worth fostering. Our agency is delicate, worthwhile, and a certain kind of achievement. Indeed, it is part of the reason we do well to praise mature forms of it, and to blame those who fail to respect it in themselves and others.⁴¹

⁴¹ Acknowledgements.

Works cited

- Arpaly, Nomy. *Unprincipled Virtue*. New York: Oxford, 2003.
- Bok, Hilary. *Freedom and Responsibility*. Princeton, N.J.: Princeton University Press, 1998.
- Borges, Jorge Luis. *Collected Fictions*. New York: Penguin, 1998.
- Bratman, Michael E. "Identification, Decision, and Treating as a Reason." *Philosophical Topics* 24, no. 2 (1996): 1-18.
- Clarke, Randolph. "Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency." In *The Oxford Handbook of Free Will*, edited by Robert Kane, 356-85. New York: Oxford, 2002.
- Dennett, Daniel. *Elbow Room*. Cambridge: MIT, 1984.
- . *Freedom Evolves*. New York: Viking, 2003.
- Doris, John. *Lack of Character*. New York: Cambridge, 2002.
- Doris, John, and Dominic Murphy. "War Crimes." Unpublished manuscript, 2004.
- Doris, John, and Stephen Stich. "As a Matter of Fact: Empirical Perspectives on Ethics." In *The Oxford Handbook of Contemporary Philosophy*, edited by Frank Jackson and Michael Smith. Oxford: Oxford, forthcoming.
- Ekstrom, L. "Free Will, Chance, and Mystery." *Philosophical Studies* 113, no. 2 (2003): 153-80.
- Ekstrom, Laura Waddell. *Free Will: A Philosophical Study*. Boulder, Colorado: Westview Press, 2000.
- Fischer, John Martin. "Free Will and Moral Responsibility." In *Oxford Handbook on Ethics*, edited by David Copp. Oxford: Oxford, forthcoming.
- . *The Metaphysics of Free Will: An Essay on Control, Aristotelian Society Series; V. 14*. Cambridge, Mass.: Blackwell, 1994.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press, 1998.
- Frankfurt, Harry. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (1971): 5-20.
- Hume, David. *A Treatise of Human Nature*. 2 ed. New York: Oxford University Press, 1978.
- Kane, Robert. "On Free Will, Responsibility and Indeterminism: Responses to Clarke, Haji, and Mele." *Philosophical Explorations* 2 (1999): 105-21.
- . "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." *Journal of Philosophy* XCVI, no. 5 (1999): 217-40.
- . *The Significance of Free Will*. Oxford: Oxford, 1996.
- Kennett, Jeanette. "Autism, Empathy and Moral Agency." *Philosophical Quarterly* 52, no. 208 (2002): 340-57.
- McKenna, Michael. "The Limits of Evil and the Role of Moral Address." *Journal of Ethics* 2, no. 2 (1998): 123-42.
- Mele, Alfred. "Kane, Luck, and the Significance of Free Will." *Philosophical Explorations* 2 (1999): 96-104.
- . "Review of the Significance of Free Will." *Journal of Philosophy* XCV, no. 11 (1998).
- . "Ultimate Responsibility and Dumb Luck." *Social Philosophy and Policy* 16 (1999): 274-93.
- Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. "Is Incompatibilism Intuitive?" *Philosophy and Phenomenological Research* (Forthcoming).
- Nelkin, Dana. "Freedom, Responsibility, and the Challenge of Situationism." *Midwest Studies in Philosophy* 29, no. 1 (2005): 181-206.

- Nichols, Shaun. "Folk Intuitions on Free Will." *Journal of Cognition and Culture* 6, no. 1 & 2 (Forthcoming in 2006).
- . "The Folk Psychology of Free Will: Fits and Starts." *Mind and Language* 19, no. 5 (2004): 473-502.
- Nichols, Shaun, and Joshua Knobe. "Moral Responsibility and Determinism: Empirical Investigations of Folk Concepts." *Unpublished manuscript* (2004).
- Ortega y Gasset, José. *Meditations on Quixote*. 1st American ed. New York: Norton, 1961.
- Pereboom, Derk. "Defending Hard Incompatibilism." *Midwest Studies in Philosophy* 29, no. 1 (2005): 228-47.
- . *Living without Free Will*. Cambridge: Cambridge, 2001.
- Sacks, Oliver W. *The Man Who Mistook His Wife for a Hat and Other Clinical Tales*. New York: Summit Books, 1985.
- Van Inwagen, Peter. "When Is the Will Free?" In *Philosophical Perspectives, 3, Philosophy of Mind and Action Theory, 1989*, edited by James E. Tomberlin. Atascadero, CA: Ridgeview, 1989.
- Vargas, Manuel. "Libertarianism and Skepticism About Free Will: Some Argument against Both." *Philosophical Topics* 32, no. 1-2 (2004 (forthcoming)).
- . "Moral Influence, Moral Responsibility." *Unpublished manuscript*.
- . "On the Importance of History for Responsible Agency." *Philosophical Studies* (forthcoming).
- . "Philosophy and the Folk: On Some Implications of Experimental Work for Philosophical Debates on Free Will." *Journal of Cognition and Culture* 6, no. 1 & 2 (forthcoming in 2006).
- . "The Trouble with Tracing." *Midwest Studies in Philosophy* 29, no. 1 (2005): 269-91.
- Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard University Press, 1994.
- Watson, Gary. "Free Agency." *Journal of Philosophy* 72, no. 8 (1975): 205-20.
- Wolf, Susan. *Freedom within Reason*. New York: Oxford University Press, 1990.
- Woolfolk, Robert L., John Doris, and John Darley. "Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility." *Cognition* (forthcoming).