

REVISIONISM ABOUT FREE WILL: A STATEMENT & DEFENSE

MANUEL VARGAS

UNIVERSITY OF SAN FRANCISCO

OCTOBER 2, 2008 | VER. 2.2B

ABSTRACT: This article summarizes and extends the moderate revisionist position I put forth in *Four Views on Free Will* and responds to objections to it from Robert Kane, John Martin Fischer, Derk Pereboom, and Michael McKenna. Among the principle topics of the article are (1) motivations for revisionism, what it is, and how it is different from compatibilism and hard incompatibilism, (2) an objection to libertarianism based on the moral costs of its current epistemic status, (3) an objection to the distinctiveness of semicompatibilism against conventional forms of compatibilism and (4) whether moderate revisionism is committed to realism about moral responsibility.

1.

Revisionism about free will is the view that an adequate philosophical account of free will requires us to jettison some aspects of our commonsense thinking about it. On this view, free will is like a host of other concepts—including scientific, moral, and conventional concepts—which we have revised to more accurately reflect our understanding of the world. Scientific cases of concept revision are plentiful. Our concepts of water, light, and temperature have undergone substantial, even radical change. But moral and conventional practices each offer a number of compelling examples of their own. Consider the fact—or, at least, I hope it is a fact—that we no longer conceive of marriage as a species of property exchange. For that matter, I take it that adultery is no longer widely thought to include any acts of pre-marital sex by the adulterer. Less salaciously, it is clear that calling David Blaine a magician does not commit us to the belief that he has supernatural powers. And (on at least some accounts), use of racial categories does not commit us to the existence of fundamental biological differences among human types. In each of these cases, our concepts of these things underwent substantial revision in light of various social, moral, and epistemic pressures. Conceptual change is thus a relatively common phenomenon, one that crops up in a broad range of cases, ranging from natural to social to artifactual kinds, across scientific, cultural, and moral categories. My contention is that like marriage, water, magicians, and adultery were, free will is ripe for a conceptual overhaul. Or, to put the point a bit differently: I think we have free will, but its nature is somewhat different that we often suppose.

Revisionism is, in at least one sense, the new kid on the free will block. What I will attempt to do here is to address some confusions that easily arise in the context of considering revisionism as an alternative to more familiar and distinguished views about free will. To that end, I will (1) sketch some motivations for revisionism about free will, (2) discuss some methodological complexities raised by revisionism, and (3) reply to some objections, especially those recently raised by Michael McKenna.¹

2.

I have maintained that an important and widespread aspect of our self-conception is that we are agents of the sort libertarianism aims to describe. Many of us are at least implicitly committed to a view that robust alternative possibilities are a pre-requisite for being apt targets of responsibility-characteristic attitudes and practices, and some of us may be committed to much stronger incompatibilist requirements (Fischer, Kane, Pereboom, & Vargas, 2007). So, with respect to characterizing an important threat to folk ontology, I have incompatibilist sympathies. Incompatibilism tends to strike many non-philosophers as obvious a description of their conceptual commitments as one can find. Nevertheless, a not undistinguished fog² of philosophers has argued that this is an error or confusion, that upon careful inspection we are not and never really were committed to incompatibilism.

Matters here are complex, both as an empirical matter about what people will assent to and as a conceptual matter concerning the proper disentanglement of our conceptual and semantic commitments. Here, I can only briefly gesture at some evidence for thinking an important thread of ordinary thinking is committed to an incompatibilist picture of agency. In particular, I will highlight three families of considerations that collectively suggest that a not insignificant portion of commonsense has incompatibilist commitments with respect to free will: philosophical, cultural, and experimental.

¹ As I will use the term, ‘free will’ is the distinctive power or ability required for moral responsibility. As I read the history of philosophy, this usage, one that links free will with moral responsibility is consistent with the bulk of at least the post-Cartesian philosophical tradition, including such figures as Spinoza, Hume, Kant, and Nietzsche. Nevertheless, I do not wish to deny that there are other conceptual roles that have been played by this term, and indeed, I will say a bit about those alternative roles later in the paper.

² My thanks to V. Alan White for coming up with the much-needed collective noun term for philosophers. It is clearly superior to James Lipton’s earlier suggestion of a *wrangle* of philosophers, which despite its not insignificant charms, strikes me as inadequate to the linguistic task, for it has failed to catch on as a linguistic convention, even among philosophers and those who despise them.

First, there is the familiar family of philosophical arguments, the most paradigmatic of which is the Consequence Argument. I tend to regard the philosophical significance of these arguments as somewhat different than is ordinarily intended by their proponents. That is, I do not take these arguments to show something about the metaphysical structure of the world. Rather, I take these arguments to constitute one kind of evidence for the conceptual commitments we have regarding free will. I see no easy or unproblematic route from the intuitions that underpin the premises of the Consequence Argument to conclusions about the metaphysics of free and responsible agency. What I do see, though, is a reason for thinking that incompatibilists have captured an important aspect of how we tend to understand the free-will-relevant notions of ability terms like ‘can’. The naturalness of incompatibilist-friendly readings of the various premises strongly suggests that even if alternative readings are possible, these alternatives do an inferior job of capturing a core or central aspect of those concepts, at least as they are implicated in freedom and responsibility.³

A second source of support for incompatibilism about folk concepts comes from (at least, but not exclusively) popular Western religious commitments. There is considerable complexity to the theological tradition in the West, but there is also a clear and influential strand of religious thinking that treats free will, understood in an incompatibilist sense, as the only thing adequate for getting God off the hook for the world’s evils. These convictions are often intertwined with an implicit commitment to a dualist metaphysics. (Dualism raises issues of its own; for example, it is not obvious that dualism of one or another sort does any better at securing non-problematic, genuine, alternative-possibilities supporting agency.) These considerations will surely seem marginal or trivial in the context of the thoroughly secular culture of contemporary analytic philosophy. But I write this in a country where the vast majority of non-philosophers (most estimates put the number at around 90%) confess to having religious beliefs. My suggestion is that popular religious convictions reflecting these influential theological commitments may be one historical source of this

³ This point can be accepted by someone who thinks, for example, that a Lewis-style counterfactual account is the best way to properly understand the involved abilities claims, and that given this, Consequence Argument-style arguments fail to show the advertised metaphysical conclusion. The matter is somewhat delicate, however. One could think that the failure of the Consequence Argument on a Lewisian account of counterfactuals shows that Lewis’ account fails to capture some important features of ordinary thinking about abilities. The significance of this result would depend on the extent to which existing folk patterns of ability talk are a constraint on the adequacy of the Lewisian story. However, if a Lewis-style account of ability claims is not so constrained, then one might advocate a revisionist conception of free will on the basis of our ordinary thinking failing to reflect the proper (Lewis-style) analysis of ability.

conceptual content, and moreover, that these religious convictions plausibly provide an ongoing narrative that reinforces those convictions.⁴

A third reason for thinking that incompatibilist convictions are central to a good deal of thinking about responsibility is rooted in ongoing experimental work on people's willingness to attribute free will and moral responsibility. Here, it is important to recognize that the data are complex and that the results point to some interesting bifurcations in the responses of experimental populations when it comes to attributions of free will, moral responsibility, praise, and blame. One thing that does seem clear, though, is that we do not get to make any sweeping claims about how all the folk are all of the time compatibilist, or all of the time incompatibilist (Nahmias, Morris, Nadelhoffer, & Turner, 2006; Nichols & Knobe, 2007). That is, the data consistently show that, depending on the experimental conditions, anywhere from 1/4 to 1/3 of the subjects will skew in some way contrary to the majority opinion. And, notably, under some conditions one gets predominantly compatibilist-favoring results, while in other conditions the results are incompatibilist-favoring. So, for example, prompts using low-emotion, abstract, or general characterizations of agents and their actions tend to generate incompatibilist responses. In contrast, prompts deploying singular, concrete cases that elicit emotional reactions tend to generate compatibilist reactions. Notably, though, when asked to describe what sort of universe is most like ours, given a choice between a deterministic one and one in which agents are non-deterministic, subjects overwhelmingly choose the indeterministic conception (Nichols & Knobe, 2007).

From a revisionist standpoint, I take all of this to be good news. Even if it turns out that we have mixed commitments —compatibilist commitments sometimes, incompatibilist commitments other times— as long as there are consistently conditions under which a not insignificant number of us have genuinely incompatibilist commitments, then it seems correct to say that some important aspect of our ordinary commitments is genuinely incompatibilist.

What should we make of the possibility of a fragmented concept? This is a complex matter⁵, one more complex than I let on in *Four Views on Free Will*, and certainly one that deserves more detailed attention that I can give it here. Part of the complexity of the matter hinges on the account of concepts one accepts. So, for example, it matters whether we accept a representationalist,

⁴ Given the diversity of theological views on these matters, one might wonder why it is that a specifically incompatibilist conception of these things might have taken root. An admittedly speculative reply: there are a variety of confusions about our agential powers that easily arise from reflection on our phenomenology of decision making, and these confusions provided the fertile soil for reception of incompatibilist and not compatibilist strands of the theological tradition.

⁵ For some of the issues involved, see (Nelkin, 2007; Doris, Knobe, & Woolfolk, 2007; Knobe & Doris, In press).

intefere[n]tial, Fregean sense, or 2-D account of concepts, and whether and how such concepts are structured. Still, it may be useful to make a few general remarks about what I have in mind by describing the folk concept as incompatibilist, and to explain why it is that even in the face of the possibility conceptual fragmentation I believe the incompatibilist is well off.

What I said in *Four Views on Free Will* was that I take it that if it turns out that there really are strands of commonsense thinking that are incompatibilist, then this is good news for incompatibilist accounts of our commonsense, and bad news for compatibilist accounts. I still think this picture is basically correct. However, a critic could reply that it is not the compatibilist's burden to knock down every conception of freedom and responsibility found in our ordinary concept: all that matters is that there is an important one that *is* part of our ordinary folk way of thinking and is compatible with determinism. Indeed, this is just what Michael McKenna has argued (this journal, ms 8).

This line of criticism raises a number of important methodological issues that can, I think, help illuminate some issues lurking below the surface of many disagreements in the free will literature. Let us start with the burden faced by the compatibilist, according to McKenna. McKenna thinks that the compatibilist's job, on the matter of describing the folk concept, is done if he or she can point to a compatibilist construal of our convictions. But such a position seems unstable, as he seems to unwilling to give the incompatibilist a similar standard of success. McKenna explicitly rejects the possibility that the incompatibilist has done his or her job by identifying one consistent strand of ordinary convictions. But on this matter there are special reasons that favor the incompatibilist's presumption, and not the view expressed by McKenna.

First, recall that by incompatibilist lights it is not enough to note that there are concepts of freedom, and perhaps notions of responsibility that intelligibly deserve those labels. As numerous incompatibilists have long acknowledged, there are plenty of senses of freedom, and perhaps of responsibility, that are compatible with determinism (Kane, 1996; Pereboom, 2001; Strawson, 1994; Smart, 1961). What is at stake, at least in the mainstream of philosophical work on free will, is the kind of freedom that is the distinctive mark of responsible agency and attendant judgments of deservingness of moralized praise and blame.

Now let us assume we're all on the same page: the sense of free will at stake is the sense of control or ability that is distinctively the mark of responsible agency. The second point is this: the folk conceptual incompatibilist ordinarily thinks that what picks out the property of free will in the world, whatever that comes to, is going to be settled by, at least in part, our thinking about free will. And, the incompatibilist ordinarily thinks, the way we figure this out is by looking at our concepts.

However, we do not have a good account of our shared concepts if we just start excising those elements that we do not like, even the ones that strike us as baroque or peculiarly demanding. Of course, if some or another commitment is just an element of one person's thinking and no one else, that commitment is poorly suited to make a claim on being our shared concept. But if you can easily get lots of people —or university undergraduates, at any rate— to vigorously agree with you under suitable conditions, then it looks like you've got a candidate for a real part of one's conceptual commitments. Alternately, we might say such a process yields data about *prima facie* contributors to semantic content. Now here is a remarkable result: the data show that, in some conditions, apparently libertarian self-descriptions run in the 90% range, and even in the least favorable conditions such reactions tend to appear in no less than 1/4 of the respondents (Nahmias et al., 2006; Nichols & Knobe, 2007). So, even if you thought we have conceptual fragmentation on free will, whether in the form of multiple concepts or merely a single inconsistent concept, it should be clear that there are conditions under which the incompatibilist strand is consistently and even overwhelmingly manifest in judgments of responsibility. So, even if there are conditions under which we really are compatibilists, the compatibilist will have hardly secured what he or she endeavors to secure if there are also a substantial number of conditions under which we are manifestly not compatibilists (Honderich, 1988; Smilansky, 2000).⁶

Here, it is important to add that what the incompatibilist is maintaining is that we are talking about the concepts of free will and moral responsibility, and not merely someone's conceptions of these things. That is, depending on your favorite view of concepts, we are talking about the broad overlap of semantic, representational, causal, or inferential structures, structures which themselves permit of an array of tokenings whose precise content is more and less elaborated. So, the incompatibilist maintains, incompatibilism is a clear feature of the concepts of free will and moral responsibility. Event-causal, non-causal, or agent causal libertarianism are arguably conceptions or theories that articulate the most promising ways of understanding the concept, in much the same way as Rawls' account of justice as fairness is a particularly promising conception of the broader shared concept of justice.⁷

⁶ Of course, one could accept this basic point and go on to argue that we would do well to excise those incompatibilist elements. So, perhaps the fragmentation of the folk requires a repair in a uniformly compatibilist way. I find this solution a compelling one, but it does not help the compatibilist on the matter of whether or not some of ascriptions of free will and moral responsibility really do, as matter of folk conceptual ontology, appeal to incompatibilist commitments.

⁷ Of course, internal to incompatibilists one might argue over whether the concept of free will has enough content to specify a particular type of libertarianism. So, for example, I take it that Tim O'Connor thinks that Kane's libertarianism

Now at this point the compatibilist might justifiably reject any number of parts about this picture. Perhaps there is an error in folk attributions of responsibility when incompatibilist judgments are manifest. Or, perhaps it is an error to suppose that such judgments reveal anything about a shared concept of free will or moral responsibility. But crucially, the compatibilist needs to offer *an argument* for why, when doing conceptual analysis, broadly speaking, we get to disregard some elements widespread in ordinary convictions, convictions that consistently show up in identifiable conditions in large numbers of people. Without an argument—and I am not aware of any having been advanced in the literature—it will surely look like the compatibilist is engaged in, well, “petty word-jugglery”; or to mix metaphors, the compatibilist is cherry-picking strands favorable to the compatibilist view and neglecting those elements of our thinking that disfavor it. These considerations also generate a defeasible reason to think that the compatibilist cannot simply retreat to a defense internal to antecedent acceptance of a compatibilist viewpoint. In the face of existing data, there is simply too much that points to ubiquitous and pervasive strands of incompatibilist intuitions in our shared thinking about responsibility. These considerations, when combined with reflections on the Western popular religious cultural heritage, and philosophical considerations of the sort delivered by the Consequence Argument and its ilk, all suggest the folk concept is rightly described as significantly incompatibilist in a way problematic for traditional compatibilists.⁸

3.

is revisionist precisely because he thinks that Kane conception of free will in fact fails to capture a genuine feature of the concept of free will. I've discussed this issue in more detail elsewhere, in (Vargas, 2005a), esp. pp. 418-419.

⁸ Here, I am ignoring some complexities concerning the principal of alternative possibilities and the numerous purported counterexamples to it, as well as the myriad of replies and counter-replies to those examples. Suffice it to say that I am not persuaded by Frankfurt-style counterexamples to the Principle of Alternative Possibilities (PAP). In brief: the only examples that are not dialectically problematic are either unpersuasive or so complex as to make clear intuitions impossible (at least for me). However, even if PAP must ultimately be abandoned I think it is likely that there is very likely a nearby principle that is best rendered in incompatibilist terms. Moreover, on the account of folk thinking I have suggested, the existence of counterexamples to a *general* principle of alternative possibilities would not show that there are *never* conditions under which folk thinking about free will and moral responsibility genuinely do invoke a principle of alternative possibilities understood in some libertarian-friendly way. But if I am wrong about both of the preceding, I still think the strength of independent arguments for incompatibilism is such that all this would show is that I've misidentified the dominant incompatibilist element in folk thinking (i.e., “leeway” instead of “source” incompatibilist intuitions), not that folk thinking does not have strongly incompatibilist elements.

Why advise that we revise? The motivating idea for my proposal of conceptual revision is two-fold. First, an adequate picture of human agency should be one that is plausible, given a broadly scientific view of the world. Unfortunately, our commonsense understanding of ourselves as free and responsible agents is not plausible in this way. I think that many libertarians, and especially Robert Kane, have done an excellent job of describing what must be true of our agency for us to make good on a widespread set of convictions about free will. The trouble is there is no principled reason to think that our agency, and in particular, the physical systems underpinning our agency, are indeed built in the way described by Kane and other libertarians. Speculative metaphysics has an unimpressive track record, and wishing our agency had indeterminism located in just the right places and none of the wrong places provides us with no reason for thinking that we have so fortuitous an alignment between phenomenology, the moments of moral concern, our causal powers, and the conditions of their collective realizers. Indeed, not only is there no evidence for the most plausible accounts offered by sober-minded libertarians, there is some positive reason to regard the proposed pictures with skepticism. Neuroscientists tend towards skepticism about free will precisely because of doubts that a non-problematic picture of indeterminism will emerge from their work.⁹

This disheartening situation leads to a second, and I think, more profound concern about our libertarian self-image, a concern about the moral status of our practices in light of the dubiousness of our self-image. To the extent to which we have social practices of blame and punishment that reflect our conceptions of agents, we have an obligation to account for the justification of those practices. However, if we are unlikely to have the kind of agency we suppose is required for the justification of those practices, it appears that we are in the morally precarious position of inflicting blame and punishment without adequate license. Unfortunately, this is precisely the position we find ourselves in. Our self-image is committed to the existence of genuine, robust alternative possibilities at nearly every moment when we are making a purportedly free choice. And, in everyday thinking, it is precisely this sort of ability to do otherwise that provides part of the justification for praise and blame. In the absence of such possibilities, goes the received view, moral

⁹ Much of the neuroscientific literature on these matters tends to be conceptually flat-footed, failing to recognize some rudimentary distinctions about deliberation, considerations, and the like. And, more often than not, neuroscientists are simply blind to the possibility of non-incompatibilist understandings of agency. So, I hesitate to lean too heavily on this literature. Still, I think it is notable that few if any neuroscientists seem sanguine about the picture Kane advocates.

praise and blame lose their justification.¹⁰ So, here's the rub: when we are honest about the evidence for our having libertarian agency, we find a dearth of evidence to support our convictions. But, given that our convictions seems to do some work in our actual practices, we folk libertarians are faced with a deeply unpalatable consequence: we advocate blame and punishment in cases where we have no good evidence for assuming that people have met the requirements for blame and punishment. Clearly, we need a better account of blame and punishment than a hope and a prayer.

This is where I take moderate revisionism to offer an alternative to (1) libertarianism, which seems to me an unwarrantedly optimistic view, (2) traditional forms of compatibilism (including, on my view, semicompatibilism), which must attribute to large numbers of people a systematic confusion between what they explicitly espouse and what they in fact believe is required of a range of moral judgments and practices, and (3) eliminativist views such as hard incompatibilism, which I take to too readily reject what can be repaired. Moderate revisionism about free will provides an alternative to each of these pictures, by showing how free and responsible agency is likely, (although at some cost to some widespread conceptions of it) and how a proper understanding of it can show that we are ordinarily licensed in our judgments of responsibility and the social practices that flow from our judgments.

A key to offering a better alternative to other, more distinguished accounts is to begin by co-opting what is best about existing accounts of free will. So, for example, I advocate understanding free will as a power or capacity bound up in a form of agency marked by sensitivity to moral considerations. However, a second key is to provide an account of the moral justification of praise and blame. As I see it, one thing that justifies the responsibility-characteristic practices and attitudes is that these things foster our moral considerations-responsive agency. Together, these two ideas generate an account that explains why moral responsibility is important, and an account of how free will is possible within a broadly scientific conception of agency in the causal world.

¹⁰ I have focused on the ideas of praise, blame, and punishment, but it is worth reminding ourselves that our conception of free will may have consequences that extend considerably further than these familiar notions, considered in themselves. I take it that, for example, attitudes towards homeless people, and drug addiction of some stripes, may reflect judgments about the freedom or responsibility of the considered agents for these conditions. If one thinks that free will is pervasive and the kind of thing that represents an unassailable ability to always transcend local motivational inputs, and we have reason to think this picture is false, this could conceivably alter how at least some people regard some conditions of homelessness and alcoholism.

One might also think that these considerations suggest that there are important strands of our thinking about agency that go beyond the alternative possibilities requirement that I have discussed, such as "source" incompatibilist intuitions, or intuitions grounded in our being at least partly non-physical substances. All of these are possibilities to which I am sympathetic. If these strands really are substantially present in our ordinary thinking, they point to further reasons for the kind of moderate revisionism I favor.

One virtue, I think, of the revisionist conception of things is that it helps to get rid of the impulse to view all non-incompatibilist accounts as “wretched subterfuges” or “quagmires of evasion.” Moderate revisionists, unlike traditional compatibilists, need not insist that ordinary, otherwise competent concept users somehow fail to realize that they do not really have incompatibilist intuitions. To the extent to which a moderately revisionist account of responsibility prescribes an ultimately compatibilist revision of commonsense, it does so in a way that recognizes the existence (but also, the limitations) of incompatibilist convictions.

Revisionism, as a methodological approach, may also help us see how to put compatibilist and incompatibilist theories into a more direct dialogue over (1) what, precisely, is at stake in our concern for free will and moral responsibility, and (2) what methods are available to us for resolving disputes about what is at stake. So, for example, moderate revisionism takes seriously the idea that we must distinguish the business of clarifying our presumptions (something akin to conceptual analysis) from the business of determining a normatively adequate conception of agency that can account why we ought to hold agents responsible. Incompatibilists have tended to focus on the first, seeking to isolate and explain the particular semantic or conceptual commitments with currency in ordinary thinking. In doing so, they have tended to highlight particularly demanding elements of our concept. In contrast, compatibilists have tended to do well in reflecting on the kinds of agency required to sustain justified normative practices. However, compatibilists have done comparatively poorly at acknowledging the concerns located in the incompatibilist strands of common sense. Each side has sought to spin out a theory of free will by attending to only one half of the story, and taking that half of the story to be the complete story. It is this implicit difference in focus—a dialectical disconnect, if you will—that partly accounts for the sensation one sometimes gets that incompatibilists and compatibilists are talking past one another, failing to understand the motivations of the other side.¹¹ When this is compounded with some methodological differences frequently found with each camp, the sense of dialectical disconnect becomes especially profound. So, for example, incompatibilist metaphysicians have often worked from a picture of metaphysics (from at least P.F. Strawson to David Lewis) that accepted a norm of descriptive metaphysics, one that sought to avoid conflict with commonsense, even at significant ontological cost. In contrast,

¹¹ Honderich and Smilansky have separately argued that compatibilism and incompatibilism both only tell part of the story about free will and moral responsibility (Honderich, 1988; Smilansky, 2000). My account is different than their respective discussions of the dialectical disconnect in an important way. They have focused on differences in the content of different conceptual strands. In addition to this difference, I have emphasized the importance of distinct foci (metaphysical vs. normative) and the attendant methodological presuppositions.

compatibilists—at least those with a foot in normative theory—have taken on board suppositions common to normative ethics. In contrast to the program of descriptive metaphysics, here there is widespread acceptance of the inevitability of abandoning some fundamental moral intuitions in the service of reflective equilibrium.

Moderate revisionism brings these disparate frameworks of concern and methodology into a unified picture. On the one hand, it permits a role for descriptive inquiry. On the other hand, it separates this issue off from the matter of prescriptive theorizing about free will and moral responsibility. In doing so, it provides a way to conceive of the relationship of commonsense to philosophical theorizing without treating philosophical theorizing as merely a project of reifying historically contingent, culturally inherited, pre-scientific prejudices.

4.

As useful as it can be to focus on methodology, we should not lose sight of the principal project, i.e., an account of how we ought to think about free will. Regardless of one's convictions about the structure and commitments of folk thinking, we can agree or dispute the accompanying positive proposal of how we ought to conceive of free will. That is, while my account's status as genuinely revisionist or not depends in part on how things sort out in the characterization of folk thinking about free will and responsibility, the success or failure of the philosophical prescription I offer is, in principle, independent of the matter of revisionism.

McKenna argues that “the tough methodological problem for Vargas is that he has to be fairly certain that the cards fall for the incompatibilist about the ordinary concept before he can proceed to cut away” (ms 7-8, in this volume). If what McKenna means is that my account is not revisionist unless I am right that the ordinary concept is incompatibilist, then I agree. And, I take it that considerations of the sort I mention in the preceding section make my construal of commonsense a plausible one. However, I fail to understand why my account is faced with a *uniquely* tough methodological problem. That is, I fail to see why I need to be any more certain than any one else with a view about commonsense thinking about free will—traditional compatibilist, libertarian, hard incompatibilist, or otherwise. If, for example, libertarians are wrong about whether folk thinking is libertarian, this is a *prima facie* problem precisely because libertarians regularly appeal to folk

thinking for justification in the libertarian metaphysics they propound.¹² Similarly, a broad swath of compatibilists would surely be dismayed for their accounts if they could become convinced that folk thinking about free will was not, in fact, compatibilist. One could defend one's compatibilism or incompatibilism at this point by appealing to the idea that in either case there is an important difference between the folk and the correct philosophical account, and that the latter rightly departs from the former. And, indeed, I think this move is an important one to make, one that makes my form of revisionism possible. But such a move would be a further move, and importantly, it would mark out a different sort of position on the relationship between folk convictions and philosophical theories than one tends to find in conventional accounts of free will and moral responsibility.

However, suppose McKenna is right. Suppose that revisionism faces a uniquely high standard of credence in its account of the folk, and suppose that this standard cannot be met. At this point, there is a very natural fallback position for my account: conditional revisionism. That is, if the chips fall a certain way, then revisionism comes alive. In the absence of suitable levels of credence about what way the chips fall, we develop the best going account we can that is prepared to walk away from various intuitions. I take it that this is all we need to justify pursuit of a theory with the various theoretical and methodological convictions that drive this account. And, of course, there are plenty of theorists who do think our concept of free will is incompatibilist, and thus, the account I offer should seem especially appealing to these folks precisely as a fallback position that does not reify groundless or metaphysically troubled aspects of common sense.

As I indicated above, I am concerned with an account of free will, understood as the control or ability condition on moral responsibility. The conceptual role of moral responsibility is one of guiding the organization, coordination, and justification of differential moralized treatment of one another. Preserving these inferential roles is what anchors any proposed revision; it is what blocks the charge of arbitrariness or the possibility that we might now revise in some direction radically disconnected from the candidate concept, as we now understand it. On this approach then, free will is roughly the distinctive control or power whose presence or absence licenses responsibility-characteristic judgments, practices, and attitudes. And, I think, the most promising account of this power is one that invokes the idea of a set of uncontroversial basic agential elements (beliefs, desires, intentions, instrumental reason) plus (1) sensitivity to specifically moral considerations and (2) the capacity to govern one's conduct in light of those considerations, i.e., moral considerations-sensitive agency.

¹² See work by Nahmias et al., 2006.

In very general terms, this account is an instance of a reason-based account of free will, of which John Martin Fischer's account is perhaps the most richly developed and extensively discussed. There are a number of important differences between my account and those favored by others, including Fischer. Here, I will not dwell on those differences. Instead, I will focus on only one aspect distinctive of my account: the special role given to how moral considerations-responsive agency structures the aims and practices of moral responsibility, and consequently, free will. As I see it, the responsibility system—that collection of responsibility-characteristic practices, attitudes, and judgments—is justified by the way it contributes to the cultivation of a particularly valuable, perhaps intrinsically valuable form of agency, i.e., our moral considerations-responsive agency. Assessments of responsibility are not merely reactions we have to one another, even if they are partly that. These assessments are part of a system whose justification and normative organization is generated by the role these assessments play in contributing to the development of moral considerations-sensitive agency.¹³ Over time, and given widespread participation in the responsibility system by agents of the relevant sort, the ubiquity of this form agency is increased and its capaciousness is enhanced. It is this fact that largely justifies the perpetuation of the general system of responsibility practices, attitudes, and judgments.

Importantly, this general, higher-order fact need not be something of which we are ever aware when we make judgments of responsibility. Instead, we have a panoply of first-order moral norms to which we appeal for the justification of praising and blaming. I get angry with you for being inconsiderate to me. I judge you responsible for that very reason. In doing so, I do not ordinarily appeal to more global normative facts. Nevertheless, those global or systemic facts are important to explaining the justification of our first order practices.

Consider an analog from the law. What makes you deserve a fine for speeding, we ordinarily judge, is that you broke the law against speeding. However, this first order judgment can be backed by a higher normative claim that never crosses your mind; one thing that can make such a law justified is the role that the cultural currency of such a law has in our society. In ordinary circumstances, when we have no need to question the justification of the law, we have no need to invoke or even consider these matters. We take the law to be justified, we see that it has been broken, and we take these things to license a judgment in favor of a fine. This is not to deny that we

¹³ We might also understand the system as “aiming at” (i.e., being justified in its operations, as a whole, to the extent to which it achieves the following results) the enhancement of agent's sensitivity to considerations where such agency already exists, as well as expanding the number of contexts in which we are considerations-sensitive agents (Vargas, in progress).

can, when engaging in the study of the law or deliberating about the proper norms governing its implementation, find matters considerably more complex. Judgments of mitigation in both law and morality are thorny matters that are pervasive in our implementation of both legal and moral judgments. At any rate, my contention is that the normative integrity of first-order responsibility characteristic practices, attitudes, and judgments is, generally, minimally secured by the role that such (usually non-consequentialist) first order judgments play in an overall moral ecology, one aimed at fostering moral considerations-sensitive agency.¹⁴

Free will, then, just is those twin capacities required for the apprehension of moral considerations and for self-governance in light of those considerations. On this account of free will, we can explain why it is important and why it has been rightly viewed as so closely tied to moral responsibility. Moreover, we can appreciate how such capacities can be common among ordinary agents. Finally, assuming we can find a way to square talk of considerations or reasons with a broadly naturalistic understanding of the universe, we can appreciate how our being free and responsible agents is compatible with various intuitive threats to free will, including determinism.

In advocating this picture of free will, I do not mean to pretend that it fully captures all the threads of ordinary thinking about it. Still, if we are willing to make some cautious departures from common sense, we can —without simply changing the subject— bring theorizing about free will more in line with the general trajectory of broadly naturalistic inquiry into human beings, one that requires us to modify our self-image in exchange for a deeper understanding of our place in nature. That’s what moderate revisionism gives us a way to achieve with our concept of free will.

¹⁴ This account is not meant to preclude the possibility of other, overlapping accounts of the justification of moral responsibility. Rather, it is meant to provide a minimal justificatory story that can be widely accepted on a range of independently plausible accounts of normative ethics. Moreover, I take it that this justificatory story gives us an analysis of what would justify the desert claims that are platitudinously invoked in responsibility ascriptions, even if (again), the correct account of the “real” normative structure of those desert claims turns out to be somewhat different than we ordinarily suppose. Thus, for example, I reject Pereboom’s conception of “basic desert” if, by that, he means a substantive account of desert that *requires* that desert is warranted by nothing more than the agent and the action, devoid of any further normative facts including social context and the broader normative significance of that act in a system of human practices. However, if Pereboom just means by “basic desert” the actual property of desert (or the best candidate properties) picked out by successful platitudinous usages of full-blooded desert attributions in the context of responsibility, then I take it that the two-tiered justificatory story I have offered just is a candidate account for the sense of basic desert involved in responsibility. In either case, my account cannot be rejected simply on the grounds that one favors an alternative conception of desert. What is needed is an argument for why, if basic term is not meant as piece of substantive stipulation, this account cannot be an account of basic desert; alternately, if ‘basic desert’ is meant to pick out a substantive stipulation, we need an argument why we should think basic desert, as opposed to the “successful platitudinous referent of fundamental desert attributions” sense of desert is the with which we are rightly concerned in discussions of responsibility. That the former has a pithier label is hardly a reason for thinking it is the sense of desert with which we are rightly concerned.

5.

By way of conclusion, I wish to briefly address three issues provoked by Michael McKenna's insightful reflections on *Four Views on Free Will* (in this journal): the nature of semicompatibilism, whether every account is really revisionist, and the matter of realism about free will and moral responsibility. Focusing on these issues can, I think, illuminate some of the commitments and attractions of the account I recommend.

Begin by considering what are we arguing about. I claimed that what is at stake in mainstream debates is the freedom required for moral responsibility. These matters are not unconnected to my claim that the terminological distinction between semicompatibilism and compatibilism obscures matters more than it clarifies them. If the freedom we are arguing about is the freedom required for moral responsibility, then Fischer and McKenna are firmly in the mainstream of traditional forms of compatibilism in that they maintain that such freedom is compatible with determinism, *irrespective of what they say about other notions of freedom ruled out by determinism*. Again—neither incompatibilist nor compatibilists have ever needed to deny that there can be *some* sense in which freedom is ruled out by determinism and some sense in which it is not. Thus, making a concession regarding some sort of freedom not required for moral responsibility is neither here nor there with respect to what most compatibilists and incompatibilists take themselves to be arguing about. So, the semicompatibilist's concession about the Consequence Argument showing that determinism rules out “an important sense of the ability to do otherwise” is not itself a notable concession in the dialectic over free will unless we go on to argue that its importance is one relevant for moral responsibility, given that this is what we are supposed to be talking about. Similar remarks are relevant for McKenna's distinguishing between “actual sequence,” or *actualist*, forms of compatibilism (such as those admirably defended by Fischer and by McKenna) and *counterfactualist* forms of compatibilism (such as classical conditional analysis compatibilists and work by more sophisticated contemporary proponents such as Michael Smith and Kadri Vihvelin). In the former case counterfactuals only play an evidential role with respect to free will. In the latter case, those counterfactuals play a constitutive role for free will.

Let me be clear: I certainly do not deny that actualist compatibilist accounts, such as Fischer's and McKenna's have made important and original contributions to our understanding of free will. Nor, am I denying that an actualist account is different from the conditional analysis part

of the compatibilist tradition. However, as far as I can tell, there is no disagreement between compatibilists and semicompatibilist, nor between actualist and counterfactualist compatibilists on whether the freedom required for moral responsibility is compatible with determinism. And, pointing to a difference between actualist and counterfactualist compatibilisms about what, precisely, is required for a positive compatibilist account of freedom does not change this fact. Indeed, it is worth remembering that actualist compatibilism has a distinguished history of its own. For example, Frankfurt and Strawson were both actualist compatibilists: neither invoked alternative possibilities as a requirement for moral responsibility, and it would be perverse to insist that their accounts are not pillars of the compatibilist tradition. Indeed, their accounts are arguably *the* defining accounts of compatibilism in the second half of the 20th century. So, although actualist compatibilism is distinct and important, I am denying both that what distinguishes semicompatibilists from compatibilists in general is the concession to incompatibilism of alternative possibilities of a sort not required for moral responsibility, *and* what seems entailed by McKenna's defense of semicompatibilism, i.e., that actualism did not have a distinguished history amongst compatibilists prior to the advent of John Fischer's rightly influential work. In this, the semi of semicompatibilism is like the semi of a semi-truck: it marks a particularly prominent instance of the kind, but an instance of the general kind nevertheless. All of which is to say that we also should not lose track of what the mainstream of the debate is about, and indeed, what compatibilism is usually taken to be about: the kind of power or agency required for being rightly blamed and praised in those characteristically moralized ways.

I do not want to pretend as though there are not other things, beyond concerns with morally responsible agency, that have a plausible claim on being an instance of the free will problem. Indeed, I strongly doubt that there is a single free will problem, or anything that adequately unifies the range of things that have at various times been considered under the label. We might, for example, be worried about *deliberative agency*, or whether we have the powers we must believe ourselves to have in order to deliberate and not have false beliefs when we deliberate (this seems to be Searle's worry, for example). Or, we might be worried about *strong agency*, or whether we have the kind of agency required for robust self-control, and perhaps characteristically human powers such as creativity and originality in decision-making (Nozick, for example, seemed to be worried about some of these issues). Or, we might be worried about *causal contributor agency*, or worried about whether we make a genuine contribution to the causal order, whatever that comes to (I take it that one strand of Nagel's concerns are located here). Moreover, these concerns need not be tied only to determinism, but could also be generated by independent worries about mechanism or physicalism or causal

explanations of human action or reductionism, and so on. In different ways, each combination of interest in some form of agency and some metaphysical or causal threat to that agency has a plausible claim on the label of *a* problem of free will. But it is also important to recognize that if it is one of these alternative instances of the free will problem about which one is theorizing, this difference ought to be highlighted. Indeed, we would do well to remember that such alternative formulations, as interesting as they may be in their own rights, are at some remove from the main of the philosophical tradition which, stretching back to at least the Epicureans, has been principally concerned with the responsible agency version of the free will problem. So, if we are to understand the ‘semi’ of semicompatibilism as flagging some concession to incompatibilists on one of these *other* versions of “the” free will problem, then all we need to clarify things is a specification of which alternative free will problem the concession was intended. In the absence of any such clarification, we should assume that what is at stake is moral responsibility and free will understood as a capacity tied to moral responsibility. If so, though, I see no reason to think actualist compatibilism or semicompatibilism is anything other than an instance (albeit a very powerful instance of) compatibilism of a general sort that includes the work of P.F. Strawson and Harry Frankfurt.

Fine, one might think. But, is revisionism vulnerable to a similar complaint? That is, is everyone really a revisionist of some or another sort, thus rendering the distinction between revisionism and nonrevisionism a specious, or at least trivial distinction? Fischer and McKenna each have flagged their willingness to revise the conditions for appropriate application of a concept, if not the concept itself. And, McKenna has gone on to ask, in light of this, whether all philosophers are in some sense revisionist. The short answer is yes, there is some sense in which all philosophers are revisionist, but it is not the sense in which I have been discussing.

There are three points relevant here. First, for familiar enough Quinean reasons, I do not find the distinction between a concept and its conditions of application a particularly compelling one, or one that gains us much in the way of explanatory utility. Second, McKenna is calling attention to a matter distinct from the focus of my concern. Recall the distinction between concepts and conceptions. Every distinct philosophical account represents a distinct conception of responsibility. Each of these is an articulation, refinement, or development of some broad commitments implicit in a more general concept that is widely shared within a given linguistic or conceptual community. What McKenna seems to be noticing is this: Conceptions are more detailed than concepts, and as such, there is a sense in which *any* philosophical account counts as a refinement or, if you insist, a revision of some folk concept. Call this sense of revision *conceptionist*

revision. However, my focus is on something distinct: *conceptual* revision. So yes, while there is a sense, the conceptionist sense, in which every philosopher is engaged in some form of revision, moderate revisionism of the sort I propose is explicitly conceptual revisionism, a thing less common than the conceptionist revision McKenna rightly suggests is common to philosophical theories.¹⁵ A third point on this matter: we should not run together at least three varieties of revisionism that might plausibly be thought to arise in the case of conceptual revision: weak, moderate, and strong. Weak revisionism is no revision to the concept itself but only to our confused beliefs we had about our concept. This is a position common to many strands of compatibilism, advocating a shift in our avowals, or self-understanding. It is a kind of revision that aims to align our second-order beliefs with our first order commitments on freedom. Thus, it is only an ersatz conceptual revision, or a minor form of conceptual revision at best. In contrast, moderate revisionism is a change in the concept itself, a change in the first order commitments, one where the concept can be rightly said to persist across the transformation. Strong revisionism is conceptual elimination or wholesale replacement. My project has been to motivate moderate revisionism about free will and moral responsibility.

Lack of methodological clarity has sometimes led some compatibilists to speak as though they were moderate revisionists about some or another aspect of responsibility and free will. Whether and how such compatibilists should be classified is no easy matter, but to the extent to which classical compatibilists have been revisionists, they seem to have been overwhelming weak revisionists (Vargas, 2005a; Vargas, 2005b). I take it that hard incompatibilists and hard determinists, when they are not fictionalists, are strong revisionists. So, in short, there is *some* sense in which compatibilists like Fischer and McKenna are revisionists: trivially, in the conceptionist sense, and less trivially but perhaps misleadingly, if we mean weak revisionists in the conceptual sense.

Finally, I wish to remark on McKenna's observation that my moderate revisionism—and, indeed, many non-eliminativist accounts of free will—is committed to a kind of realism. This is a matter I do not directly address in *Four Views on Free Will*, but it is one I have tried to address in other places (Vargas, 2005a; Vargas, 2004). Briefly, here is how I take things to stack up.

First, I am hesitant to agree that my account, or others like it, are indeed committed to free will realism, simply because philosophical discourse over realism is fraught with disagreement about

¹⁵ Elsewhere, I have offered a taxonomy of a wide range of possible varieties of revisionism, and distinguished my own brand of moderate revisionism from several alternative varieties, including several possibilities I do not mention here (Vargas, 2005a). For usefully related discussions, see also recent work by Shaun Nichols (2007)

what, exactly, realism comes to. So, for example, in various places I have helped myself to characterizing talk of responsibility and free will as truth functional. This is a matter of convenience. Given that *moral responsibility* —and by extension, perhaps, *free will*— count as moral predicates, everything I say should be translatable into any adequate non-cognitivist account, if noncognitivism about moral terms is correct. So, if noncognitivism is true and my account were translated into a discourse that makes explicit sense of that architecture of meaning, it would be a bit strange to insist that the result would be a theory committed to problematic realism about moral responsibility.¹⁶

Second, even if we accept the idea that there must be a commitment to *something* realist here, I have argued that it may make no practical difference if what we refer to is not free will in some strict sense, just so long as it is otherwise a twin of free will with respect to its social, conceptual, and normative roles (Vargas 2005a, Vargas 2004). Against those species of anti-realist views that I have called *property error-theoretic views* (e.g., Pereboom's), I have maintained that they face a special explanatory burden. The burden is this: if we can demonstrate the obtaining of something else that does virtually all the same work as some purportedly non-existing *X*, the error-theorist must account for why it matters that he or she has an argument for the non-existence of *X*.

Let me explain. Even if it turns out that the positive account I offer is not an account of free will, for perhaps Pereboom is right that free will, strictly speaking, does not exist, it does not follow that some very close analog, free will*, also fails to obtain. Indeed, on this account, we have an excellent candidate for free will*: namely moral considerations-sensitive capacities. Now here's clincher: (1) if free will* does everything we want of free will (e.g., it preserves the bulk of inferences licensed by the concept, its presence would warrant characteristic moral evaluations and its absence warrants the suspension of those reactions— recall the bit I noted earlier about these things being the anchors for any proposed revision), *and* (2) it has the virtue of existing, then (3) it seems like we would do well to abandon any concern for free will and move on to the theoretically and existentially superior business of keeping track of free will* (Vargas, 2004). And, I submit, that is exactly what my notion of moral considerations-sensitive agency permits us to pursue in the face of non-existence claims.

¹⁶ How this rendering of my account would intersect with talk on the freedom or ability condition of responsibility, i.e., free will, is a complicated matter. One might think responsibility talk is noncognitive even if, for example, it appeals to a (perhaps) truth-functional concept of free will. And, of course, there is the further matter of how we should understand the semantics of free will, something that on my account may yet turn out to be internalist or externalist in its semantics (Vargas 2005a).

So, maybe anti-realism about free will is warranted. If so, though, it is not clear that such a possibility has practical consequences, given that we have a good case for realism about free will*, that is, an account of something that can do virtually all of the important work of a theory of free will.

Now, as things stand, I do not think there is good reason to favor an account of the semantics and reference of free will that would result in an antirealist interpretation of free will. Elsewhere I have argued that we are better off adopting something like a Lewis-style functional role specification of the denotation of the term, permitting the possibility of richly incompatibilist connotational content that does not settle the matter of successful reference (Vargas, 2004). Instead, I have suggested, reference here is plausibly settled by whatever instantiated property (or properties) does the best job of making true the vast majority of platitudinous ascriptions in which free will is implicitly or explicitly invoked, while preserving characteristic patterns of inference and licensing of normative judgments.¹⁷ And, I believe, my account is an account of exactly the sort of thing plays precisely this functional role.

So, I have indeed directly argued against anti-realist accounts of free will. However, even if antirealism about free will remains plausible, its consequences are comparatively minimal. What we would find, I submit, is that antirealism about free will to be as worrisome as, say, antirealism about caloric turned out to be. Caloric was thought to be a fluid or gas that flowed from hotter to colder bodies. Once we had a viable theory of heat, however, a theory that did all the work that made us concerned about caloric in the first place, it was not really worrisome to concede the nonexistence of caloric. So, although we cannot rule out the possibility of antirealism about free will, I do think the heat is on antirealists to explain why we should be concerned about the possibility antirealism in the face of viable alternatives that do all the principal work that drove us to worry about this sense of free will in the first place.¹⁸

¹⁷ One virtue of my account is that it just is the application of a much broader story about how to think about the rational constraints on language use. It is not as though we suddenly ignore general rules about concept usage or the achievements we've made in understanding philosophical issues about language. Nor do we commit ourselves to a picture where we treat the concept as unique and sacrosanct, off in its own insulated space of conceptual and logical constraints, detached from the world and our purposes. I have said more about this picture in (Vargas, 2004), although there I do not there emphasize the preservation of characteristic patterns of inference and licensing of normative judgments, partly because I was assuming (rightly or wrongly) that such things would automatically embedded in the structure of platitudinous ascriptions of freedom and responsibility.

¹⁸ I wish to thank Joseph Campbell for organizing the APA session that led to this paper. Thanks also to my co-authors of *Four Views on Free Will*, as well as Richard Holton, Michael McKenna, and the indefatigable Dan Speak for fruitful discussions and exchanges about the material in this paper. I also wish to acknowledge the feedback from audience members at the 2008 Pacific Division Meeting of the APA. And, I am grateful for the generous support of the Radcliffe Institute for Advanced Study at Harvard, where I completed work on this paper.

BIBLIOGRAPHY

- Doris, J., Knobe, J., & Woolfolk, R. L. (2007). Variantism about Moral Responsibility. *Philosophical Perspectives*, 21(1), 183-121.
- Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M. (2007). *Four Views on Free Will*. Malden, MA: Blackwell.
- Honderich, T. (1988). *A Theory of Determinism*. New York: Oxford.
- Kane, R. (1996). *The Significance of Free Will*. Oxford: Oxford.
- Knobe, J. & Doris, J. (In press). Strawsonian Variations: Folk Morality and the Search for a Unified Theory. *The Handbook of Moral Psychology*. Oxford: Oxford University Press.
- McKenna, M. Compatibilism and Desert: Critical Comments on *Four Views on Free Will*. 2008 *Pacific Division Meeting of the American Philosophical Association*.
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research*, 73(1), 28-53.
- Nelkin, D. (2007). Do We Have a Coherent Set of Intuitions About Moral Responsibility? *Midwest Studies in Philosophy*, 31, 243-259.
- Nichols, S. (2007). After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes. *Philosophical Perspectives* 21, 405-428.
- Nichols, S. & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Nous*, 41(4), 663-685.
- Pereboom, D. (2001). *Living Without Free Will*. Cambridge: Cambridge.
- Smart, J. J. C. (1961). Free Will, Praise, and Blame. *Mind*, 70, 291-306.
- Smilansky, S. (2000). *Free Will and Illusion*. New York: Clarendon.
- Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies*, 75, 5-24.
- Vargas, M. (2005a). The Revisionist's Guide to Responsibility. *Philosophical Studies*, 125(3), 399-429.
- Vargas, M. (2005b). Compatibilism evolves? On some varieties of Dennett worth wanting. *Metaphilosophy*, 36(4), 460-475.
- Vargas, M. (in progress). Building A Better Beast. *Unpublished manuscript*.
- Vargas, M. (2004). Libertarianism and Skepticism about Free Will: Some Arguments Against Both. *Philosophical Topics*, 32(1&2), 403-426.