REVISIONIST ACCOUNTS OF FREE WILL: ORIGINS, VARIETIES, AND CHALLENGES

Forthcoming in The Oxford Handbook of Free Will, *2e*

Manuel Vargas

Last updated: Feb. 11, 2010

The present chapter is concerned with revisionism about free will. It begins by offering a new characterization of revisionist accounts and the way such accounts fit (or do not) in the familiar framework of compatibilism and incompatibilism. It then traces some of the recent history of the development of revisionist accounts, and concludes by remarking on some challenges for them.

1. Revisionism and eliminativism

We change our mind about things all the time. Sometimes those changes do not involve any important or deeply held beliefs. If I think it is not going to rain today, and it starts raining as I leave my house, I'll simply update my beliefs to reflect my conviction that it is indeed a rainy day. Other changes, however, can demand more subtle readjustments in our web of beliefs.

Consider a theist who thinks that the content of morality is and must be determined by the decrees of God. Now suppose that this person —we'll call him Friedrich— starts to doubt whether or not God exists. He eventually becomes an atheist. Consider what Friedrich might now say about morality, given his newfound atheism.

Friedrich could reject morality. Given his conviction that morality requires God, he might conclude that morality is a sham.

Having rejected morality, Friedrich might go on to think that we have good reason to act as though morality were not a sham. Or, he might not. Perhaps he would think it is a good thing if we all realized the illusory nature of morality. Practical considerations aside, however, if Friedrich followed this line of thought to its conclusion, he would hold that morality ought to be eliminated from the catalog of what truly exists in our world. Call this view *eliminativism* about morality.

Eliminativism is not Friedrich's only option. Instead of thinking that morality should be rejected, he could think that what needs rejection is his conviction that morality requires God's existence. This might be an especially tempting option if he came to believe that morality's distinctive judgments and practices could be grounded in practical reason (or sentiment, or human functioning, or ideal social arrangements with which no one could reasonably disagree, or . . .). Of course, Friedrich would have to acknowledge that atheism changed things for him. He would not regard morality in quite the same way as he had before. Nevertheless, if he reasoned in this way he would likely insist that morality should be included in the catalog of what truly exists in our world. Notice, though, that if Friedrich's former theistic conception of the foundations of morality was widespread, his newfound convictions about morality would put him at some distance from conventional ways of thinking about it. To mark this fact, we can call this view *revisionism* about morality.

Eliminativism and revisionism are not positions limited to morality. When the biological notion of race came under widespread criticism, there were some who thought that the defects of the biological notion showed that there were no races. Others thought that the right view was to think that although races exist, the nature of race was not biological. The former reaction was eliminativist and the latter revisionist. When cognitive science and neuroscience put pressure on standard taxonomies of mental terms,

there were those who took this to show that eliminitivism about folk psychology was the correct view. Others thought that, at best, the threats from the sciences of the mind showed that we needed to revise some of our folk psychological categories. So, eliminativism and revisionism constitute options in a variety of theoretical domains.

2. Revisionism and the compatibility debate

For at least the past thirty years, the standard way of characterizing the landscape of the free will debate is by reference to the compatibility question— *is free will compatible with the thesis of determinism*? If one says yes, then one is a *compatibilist*. Since the vast majority of compatibilists have gone on to maintain that the requirements of free will are satisfied, most take it that a commitment to compatibilism brings with it a conviction that we at least sometimes have free will. If, however, one rejects the compatibility of free will and the thesis of determinism, then one is an *incompatibilist*. There are two main incompatibilist options. Either one is a *libertarian*, holding that we possess free will, at least sometimes. Or, one is an eliminativist, holding that we lack free will because, for example, determinism is true.[1]

The recent free will debate has been substantially shaped by both the centrality of the compatibility question and a particular picture of methodology in metaphysics. On the broadly "intuitionist" model of metaphysics, we generate a theory of free will via reflection on our concepts as we find them, and we test

---

1 The view that the thesis of determinism is true and that we therefore lack free will is usually called *hard determinism*. Partly in light of how quantum mechanics is ordinarily understood, holding that we lack free will whether or not determinism is true has been a more common view. *Hard incompatibilism* is Derk Pereboom's helpful label for views of this sort.

proposals by checking to make sure they do not run afoul of our intuitions about cases.[2] We may tolerate some counterintuitiveness when it is in the service of a more powerful intuition. Arguments for eliminativism usually work in this fashion, generalizing from our judgments about cases, or some purportedly powerful intuition about the necessity of some condition that we (ultimately) learn is impossible to satisfy (e.g., Pereboom in Fischer et al. 2007). Throughout, the governing presupposition is that our metaphysics of free will can be read off of our beliefs about free will, our intuitions about cases and principles, and what these imply.

There is a fundamental worry about this approach, which we might label *the Sellarsian worry*. The worry is that the prevailing methodological approach runs the risk of conflating what Sellars called our "manifest image" of the world (in this case, our standing commitments concerning free will), and the "scientific image," (here, the prescriptive account of how we ought to think about free will, all things considered). For example, incompatibilism about our ordinary concept of free will does not entail incompatibilism about how we should think about free will. This might seem like a trivial thing to keep track of. After all, one might say, what we were interested in all along was the nature of free will, not our standing beliefs or concept.[3] Perhaps.

Notice, however, that this concession forces us to reconsider whether many of the familiar arguments in the literature—the Consequence Argument, Frankfurt cases, the Four Case Argument, the Basic Argument, and so on—show us anything at all about the nature of free will, as opposed to our manifest image of it. Standard

---

2 Perhaps the classic model of this approach is Strawson's (1992) project of "descriptive metaphysics" or "concept-mapping." Compare Lewis (1973). I've discussed these matters in greater detail in Vargas (forthcoming*c*).

3 Kevin Timpe has raised this worry.

arguments are ordinarily taken to show something about the nature of free will. However, once we allow that our received commitments (our intuitions, our concept, our collection of beliefs) about free will and its nature can come apart, it is not clear why we should suppose that those familiar arguments show us anything about free will. What exactly is it that moves us from reflections about our manifest image to conclusions about the metaphysics of free will? Perhaps such arguments can be developed, supplemented, or defended in some way that makes clear the vehicle from our intuitions or received commitments about particular cases or principles to knowledge of the metaphysics of free will. However, what the Sellarsian worry suggests is that, at the very least, a good deal more is required before we can allow that it is obvious that these arguments teach us anything about free will itself.

On this picture, the conventional framework of compatibility leaves out something important: the distinction between our manifest and scientific images. If we assume these things come to the same thing, the possibility of revisionism is removed from the start. Once we allow that these things might be different, we open up space for a revisionist account.

We can begin to characterize that space if we start with the idea that whether an account is revisionist with respect to something depends in part on our ordinary commitments about that thing. If, for example, no one was ever really committed to a divine command theory of morality, than a proposal for the non-divine foundations of morality would not automatically count as revisionist. Similarly, for a theory of free will to be revisionist, it must propose an account that departs from our ordinary commitments about free will. So, we can distinguish between a *diagnostic* (or descriptive) account of free will and a *prescriptive* account of free will. The former aims to characterize the kinds of commitments ordinarily had about free will, whereas the latter is a proposal for the commitments that, all

things considered, we ought to have.[4]

We can now fix some terminology. *Conventional* accounts entail consistency between prescription and diagnosis.[5] Consistency between diagnosis and prescription does not settle whether free will is, all things considered, compatible with determinism.[6] Conventional compatibilism is the view that, as matter of *both* our folk concept and our best prescriptive theory, free will is compatible with the thesis of determinism. Conventional libertarianism is the view that, as a matter of both our folk concept and our best prescriptive theory, free will is incompatible with the thesis of determinism and we have free will. *Revisionist* views are those on which the proposed prescriptive account conflicts with the diagnostic account. Prescriptive accounts may be compatibilist or incompatibilist: there can be revisionist compatibilists and revisionists incompatibilists. What sets revisionist accounts apart from their conventional counterparts is the contention that we

---

4 Because of the 'descriptive/prescriptive' distinction has connotations in metaethics, I prefer to use the label 'diagnostic'. That said, I have no objection to using 'descriptive' rather than 'diagnostic', so long as we are careful to distinguish between accounts that endeavor to characterize our (perhaps latent) commonsense views and proposals for how, all things considered, we should characterize the matter. See, also, Nichols' (2006) tripartite division of theoretical aims. I've avoided Sellars' own terminology, for it may suggest that any prescriptive account will necessarily be "scientific," whatever that comes to.

5 I recognize that there is some infelicity here, as 'conventional' can also describe things whose truth conditions appeal to conventions. I mean conventional in the "customary" or "ordinary" sense of the word.

6 By distinguishing the conceptual terrain in terms of the relationship of free will to determinism, I do not mean to preclude the possibility that there are other threats (e.g., naturalism, divine foreknowledge, social psychology, etc.) to free will or moral responsibility.

should abandon some commitments that constitute our ordinary way of thinking about free will.

Notice that if we maintain that *any* difference between diagnosis and prescription is sufficient for revision, then revisionism threatens to become an uninteresting category. After all, many conventional accounts of *X* invoke commitments that are not a part of ordinary beliefs about *X*.[7] For example, a particular deontological account of morality might stipulate the requirement that moral truths about some considered course of action are to be determined by testing the potential maxims under which one acts against the demand to treat people as more than mere means to some end. Even if it turned out that ordinary beliefs about morality did not involve a commitment to maxims, beliefs about treating others as mere means, and so on, it would be strange to insist that such commitments constitute a revisionist account of morality—*unless such elements were in conflict with ordinary moral commitments*. This suggests revisionist accounts are best construed as those on which the prescription includes commitments that are not merely absent from common sense, but that are in conflict with it. On this way of labeling the difference, an account of free will that (say) invoked the idea of neurological magnification of quantum indeterminacies would not be revisionist unless the idea of such magnification was inconsistent with commonsense commitments about free will.

Revisionist accounts thus recommend a positive account of free will that not only departs from, but also conflicts with, aspects of commonsense thinking about free will. It is why revisionist accounts are at odds with conventional accounts of free will. This may suggest that revisionism and eliminativism are kin, from the

---

7 This is an objection that McKenna (2009) rightly made against my earlier characterizations of revisionism. The present account, which follows, is intended to supersede my previous response (Vargas 2009).

standpoint of types of theories. Previously, I said as much (Vargas 2005a). I now regard this as a mistake. Revisionism and conventionalism are views about the relationship of our prescriptive theories of the nature of free will with our pre-philosophical views about these things. Eliminativism is a position that denies the existence of free will, regardless of whether our best theory of it is revisionist or conventional. Of course, most revisionist and conventional accounts of free will are intended as success-theories, that is, committed to the view that we have the thing in question (free will). Still, revisionist and conventional accounts need not reject eliminativism, even if they ordinarily do.[8]

3. Varieties of revisionism

So, revisionism is distinct from conventional compatibilist and incompatibilist accounts. Here, I want to say a bit about revisionist options and nearby alternatives that might appear to be revisionist but are not.

Recall Friedrich and his reconsideration of the divine command theory of morality. Let us suppose that Friedrich comes to regard eliminativism with suspicion and begins to weigh up his non-eliminativist possibilities. One thing he might conclude is that his initial diagnosis of his convictions was in error. He could think he misdiagnosed his actual beliefs about morality. Perhaps he concludes that even though he might have explicitly avowed a divine

8 On this way of fixing terminology, one could be an eliminativist about free will and either a conventionalist or a revisionist. Eliminativist revisionism would occur if, for example, one thought that the best theory of free will conflicted with the commitments of common sense, but that even on our best revisionist theory we still lacked free will. Notice, too, someone who thinks that free will is compatible with determinism but that we lack it for other reasons (e.g., Neil Levy), could be a conventional compatibilist and an eliminativist.

command theory of morality, he was not actually committed to it, and it was his acceptance of atheism that helped him see that his avowals were not reflective of his genuine commitments. Whatever the plausibility of this maneuver, it is not a revisionist one. At best, it is an epicycle on a conventional account. It does not hold that there is a conflict between the proper diagnosis and the proper prescription, as revisionists do. It simply holds that the initial diagnosis was in error. Call this possibility *diagnostic correction*.

The properly revisionist route begins not from amending the diagnostic account. Instead, it proceeds from accepting the diagnosis that one's actual commitments contain an important error. In Friedrich's case, this would mean that he now regards his prior commitment to divine command theory as an error, something to be abandoned. It is hard to say what sort of thing it is that he now thinks he must abandon (a belief? the concept? a conception?), but for present purposes whatever it is that he abandons will be significant in either of two ways.

First, he could be giving up mental content that plays no role in fixing the referent of the term, something we can label *connotational content*. That is, Friedrich's beliefs about divine command theory could be epiphenomenal to reference. They would be, roughly, beliefs he associates with free will, but they would do no substantive work in designating some property in the world. Indeed, Friedrich might take the fact of some alternative proposal for understanding morality (e.g., Kantian, contractualist, consequentialist, virtue theoretic, etc.) to count as evidence that his divine command theory beliefs were only connotational, despite his prior belief that they were essential to reference-determination.[9] If so, then Friedrich is a

9 Notice that those beliefs about his theoretical commitments count among connotational content; his prior beliefs about whether those other beliefs were reference-fixing does no work in fixing reference.

connotational revisionist.

A second possibility: Friedrich could be giving up some reference-fixing content. If so, one might think, Friedrich is really committed to eliminativism. After all, he is advocating giving up commitments that do work in referring to morality. Not so fast, though. Suppose that Friedrich thought that there is some nearby property, very much like morality, which exists and to which we could refer. Suppose further that it occurs in all the places where we customarily attempt to refer to morality, and its presence or absence licenses the practices and attitudes characteristic of morality. So, for example, suppose the not-really-morality-but-pretty-close-to-it account of contractualism explained the truth of various not-really-morality-but-pretty-close-to-it claims, claims that license morality-characteristic reactions, attitudes, and practices. Friedrich might eschew eliminativism precisely because he thinks that we can re-anchor reference to morality without violating any of its important features. After all, this new stuff preserves the primary inferential roles we take to organize our beliefs about morality and it regiments our practices and characteristic attitudes in familiar ways. Moreover, the new stuff weighs in our deliberation in just the same way that morality-in-the-strict-and-unrevised sense does, and in general, it preserves morality's characteristic normative import. If all this were true, Friedrich might think that a revision of reference, *denotational revision*, is warranted. In doing so, he would be advocating that we change the topic, in some sense, but in a way that respects the fundamental work of the concept (Vargas forthcoming*a*). Indeed, if all this were true, Friedrich-the-denotational-revisionist would think the onus is on eliminativists about morality to explain why we should care about the loss of morality in the old, impossible sense. After all, this Friedrich would say, we have a plethora of non-divine command accounts that preserve the import of morality and adequately explain the practices, attitudes, and inferences that we

take to make up our moral lives and discourse.[10]

These three possibilities—diagnostic correction, connotational revision, and denotational revision— map on to the free will debate in relatively straightforward ways.

The best-known instance of theorists who championed a diagnostic correction may be classical conditional analysis compatibilists. Such compatibilists were sometimes prepared to grant that people might have *said* that they believed in a "categorical" or non-conditional conceptions of 'can', but what they failed to see was that any such notions in fact bottomed out in a conditionalized ability. Similarly, contemporary compatibilists and their near relatives will sometimes grant that people have incompatibilist intuitions about free will. However, these theorists maintain that careful philosophical reflection shows that such characterizations are at root failures to appreciate one's more modest actual commitments.[11] So, on a diagnostic correction, what needs correction is not our commitments, but our appraisal of them.

---

10 In opting for revisionism about morality, Friedrich may be unsure about whether he is committed to connotational revision or denotational revision. As I noted above, ordinarily available evidence may insufficient to tell us which characterization is true. Consequently, Friedrich might accept a kind of *semantic agnosticism*, if he were unsure about how to parse questions of whether his old, problematic beliefs about divine command theory were connotational or denotation. For him, the important thing is that, either way, he accepts a revisionist account.

11 Fischer sometimes suggests this, highlighting the somewhat revisionist elements in his work while at the same time holding that reflection on philosophical thought experiments show that (at least regarding moral responsibility) we do not really have such commitments. For some discussion, see Fischer et al. (2007, chapter 8). I confess that I am now less certain how Fischer intends his account, although I welcome the possibility that he is revisionist about moral responsibility, even if it strikes me as dubious that he is revisionist about free will.

Connotational revisionists about free will hold that we need to expunge aspects of folk thinking about free will, but that in doing so we do not disrupt reference. On this view, we have been talking (successfully, let us presume) about free will all along, even if we had erroneous beliefs about it.

Dennotational revisionists hold that we need to expunge aspects of folk thinking about free will, and that in doing so we will re-anchor the referent for 'free will'. What makes a denotational revision possible is that we can re-anchor our talk of free will on some property whose existence is in most or even all of the places we used to refer to free will (in the pre-revised way), and the fact that its presence or absence warrants the typical inferences, reactions, and social practices that are characteristic of free will.[12]

In prior work, I have distinguished between weak, moderate, and strong revisionism. Weak revisionism is what I am here calling diagnostic correction. Strong revisionism is eliminativism. Moderate revisionism was ambiguous between connotational and denotational revision. I am now unhappy about the weak/moderate/strong distinction for several reasons. First, moderate revisionism's ambiguity between connotational and denotational revisionisms invited confusion. Second, weak and strong revisionism are not ordinarily revisionism at all. Weak revisionism (diagnostic correction) is a conventionalist's admission that he or she mischaracterized our commonsense views about free will. Strong

---

12 Pereboom (2009) has argued that the characterization of revisionism depends partly on the picture of concepts involved. Notice, though, that revisionism (whether denotational or connotational) is largely neutral about the characterization of concepts. The salient issue is whether what is getting revised is connotational or denotational in its significance. So long as we can render this distinction on your favorite view of concepts, we have all that we need to capture the main contours of what is significant about revisionist accounts.

revisionism (eliminativism) is a view that holds that we should reject the existence of free will, irrespective of whether our best account of free will's nature is at odds with our folk conception of it. So, I now propose that we regiment terminology in the way I have suggested here, reserving 'revisionist' for those theories that are committed to either denotational or connotational revision.

4. The development of revisionist views

In attempting to trace the development of revisionist accounts since the early 1960s, we immediately face two difficulties. First, prior to the early 2000s, philosophers rarely if ever described their accounts as revisionist, even when it contained revisionist elements. Second, many of the accounts that could be plausibly characterized as revisionist have been subject to contested descriptions on just this issue. For example, philosophers sometimes associate revisionism about free will with Daniel Dennett's slogan about the varieties of free will "worth wanting." However, Dennett's own understanding of his account is hard to square with this perception, for he maintains that the varieties of free will worth wanting are not distinct from what we in fact want.[13]

A different puzzle with another prominent candidate for an early revisionist arises with Jonathan Bennett's penetrating discussion of Strawson's "Freedom and Resentment" (2003). Bennett (1980) reads Strawson as a revisionist. However, his interpretation of Strawson is difficult to square with two aspects of Strawson's work. First, there is Strawson's famous rejection of "revisionary" metaphysics in favor of a descriptive metaphysics wherein one limits one's philosophical theorizing to explicating the content and implicit commitments of our conceptual framework as

---

13 Dennett (2003) explicitly distinguishes his view from a revisionist account, which he associates with White (1991). For discussion, see Vargas (2005b).

we find it. Second, there is Strawson's explicit claim to describe "what we mean" and "all we mean" by free will (91). There is, of course, the matter of whether and to what degree compatibilists have taken seriously such a constraint on their theorizing (Singer 2002, Vargas 2004). It is open to one to claim, as Frank Jackson (1998) has done, that whatever it is that compatibilists take themselves to be doing, they *ought* to understand themselves as doing something like Quinean paraphrasing, i.e., a limited form of changing the subject to something that is near enough to the folk conception "to be regarded as a natural extension of it"(44). Indeed, this has been advocated by a number of revisionists. However, unless we understand compatibilists as generally committed to accounts that cohere with our concepts and practices as we find them, then we could make no sense of Hume's familiar claim that "the whole dispute . . . has been hitherto merely verbal." For that matter, it would be equally hard to make sense of Kant's charge that compatibilists are engaged in "a wretched subterfuge."[14]

I am inclined to think that J.J.C. Smart's (1961) account is the clearest forerunner to contemporary revisionist accounts. Smart argues that the "metaphysical" or "libertarian" theory of free will is incoherent. He also maintains that much of the justificatory work for which libertarian or metaphysical views of free will are invoked is ultimately unnecessary. So, for example, he claims that "threats and promises, punishments and rewards, the ascription of responsibility, and the non-ascription of responsibility" have justification that is compatible with belief in determinism (302). However, he concedes that moral blame, as it is ordinarily deployed, does presume the libertarianism's incoherent metaphysics.

---

14 Compare Singer (2002): "Compatibilists typically present their view as descriptive rather than as revisionary, and often criticize libertarianism for its distorted description of ordinary beliefs" (28).

Consequently, he recommends that we restrict ourselves to a notion of "dispraise" that is tantamount to a kind of grading, or a statement of "what people are like" (304).

It is not clear to what extent Smart takes himself to be offering an alternative (compatibilist) account of free will, or whether he holds that 'free will' is a term of confused metaphysics. Still, the suggested emendation of those portions of our practices concerned with blame anticipates some elements of contemporary revisionist accounts. As he sees it, there are portions of our ordinary beliefs about freedom and blame that stand in need of repair, but such repairs do not require wholesale abandonment of familiar distinctions concerning agency, abilities, and the bulk of their connection to moral evaluations.

Although Smart's work engendered some fruitful discussions, it did not spark a surge in revisionist work. However, more recently there has been a marked increase in accounts of a broadly revisionist extraction. I believe it is traceable to two independent sources, both outside the literature on free will.

One of these sources is externalist theories of content and reference in the philosophy of language. In a 1996 article, Mark Heller argued that Putnamian insights offered the possibility of a new kind of compatibilist account. Putnam's idea was that conceptual content was oftentimes insufficient to settle the matter of the real essence of things. On this view, conceptual content alone is insufficient for us to learn that water is H_2O. Heller argued for the possibility of a kind of compatibilism that exploits this idea, one where the compatibilist grants that there is no satisfactory compatibilist analysis of the concept of free action, but that compatibilism nevertheless captures the correct essence of free action. Again, there was comparatively little direct uptake of this proposal, at least at first. Nevertheless, Heller identified a conceptual possibility that became important for work that followed

not long after.

A second source of recent revisionist theorizing is rooted in the dominant methodology of moral theory. Over most of the past 100 years of ethical theorizing, intuitions have been widely regarded as important, but hardly sacrosanct. Most ethicists accept that it is impossible to capture all the divergent intuitions ordinary people have about ethical issues. The governing idea has been the sort of thing Rawls described as "the method of reflective equilibrium." Normative theories are to be generated by working back and forth between considered judgments (or intuitions), the principles we take to explain them, and the considerations that speak in favor of those judgments and principles. It is a picture on which revisionary outcomes are taken for granted, and the project of a "descriptive ethics" along the lines of "descriptive metaphysics" is widely regarded as a non-starter.

It is no surprise, then, that among accounts that have been concerned with free will as it arises in connection with moral responsibility, there have been intimations of revisionism. So, for example, R. Jay Wallace (1994) and T.M. Scanlon (1988) have both offered accounts of the capacities required for moral responsibility, and they have noted that their accounts are potentially revisionist with respect to retributivism. It is, of course, a further step to conclude that such revisionism is rooted in free will. Still, in much of the literature a shift between talk of those capacities that underwrite moral responsibility to talk of free will is ubiquitous. Numerous philosophers have claimed that free will just is the "freedom" or "control" condition on moral responsibility (Pereboom 2005; Fischer et al. 2007; Mele 2009). So, there is some reason to interpret Wallace and Scanlon as revisionist about free will, even though neither labels himself as such.

These threads of influence from the philosophy of language and moral theory came together in the early 2000s in the work of

three philosophers each of whom independently proposed and explicitly identified a 'revisionist' account of free will or moral responsibility (Hurley 2000; Singer 2002; Vargas 2001).

Susan Hurley's (2000) articulation of a revisionist position grew from her dissatisfaction with an argument of Galen Strawson's for the impossibility of moral responsibility. In its most basic form, Strawson's (1994) argument is this: (1) Nothing can cause itself, (2) In order to be morally responsible for one's actions, one would need to the cause of oneself, and (3) therefore no one is morally responsible for any actions.

Hurley focuses on moral responsibility (as does Strawson) and takes issue with Strawson's argument on two grounds.[15] First, she contends that it is improper to simply assume that impossible properties can be essential, as Galen Strawson does. Second, she argues that once we consider going theories of reference, it is implausible to contend, as Strawson does, that what is at stake is an impossible essence.

Crucial to Hurley's (2000; 2003) approach is the idea that on any of the main theories of reference, Strawson's argument does not go through. Take a picture on which the essence of a term is that which must be understood to know what it means. Now suppose one thought that some impossible essence is reference fixing because what fixes reference is armchair reflection on the meaning of terms. First, this view seems to require a commitment to analytic truths that many contemporary philosophers reject as untenable. Second, it runs afoul of the fact of disagreement about the meaning of 'moral responsibility'. In the face of disagreement about the

---

15 Hurley (2000) notes that she and her interlocutor take seriously the idea that their disagreement about moral responsibility is ultimately intertwined in matters of freedom, but her argument for revisionism proceeds explicitly in terms of moral responsibility (notes 43 & 45).

essence, it is especially dubious to think that that what we are discerning from the armchair is an essence. So, she thinks, this view in defense of impossible reference should be rejected.

Now suppose one thinks that reference is fixed by contexts of actual use. Here, it is difficult to see how an impossible essence could be the thing picked out by contexts of actual use. As Hurley (2000) puts it, impossible essences lack adequate "explanatory depth" (245). That is, because such essences cannot obtain, it is precisely for that reason that it is implausible that such things would be picked out by contexts of actual use to explain phenomena. Similar remarks hold in the case of functional term accounts of reference: a degree of explanatory depth is required for something to be a candidate realizer of 'moral responsibility' and essentially impossible essences can play no role in explaining phenomena. Consequently, if there are other realizers available, the most natural thing to say about the proposed impossible essence is that the error was in identifying it as an essence.[16]

A powerful feature of Hurley's account is that it relies on considerations that derive their plausibility independently of any particulars in the free will debate. Assuming Hurley is right about what is entailed by contemporary theories of reference, this leaves the "impossible essence" eliminativist about moral responsibility with an unappealing dilemma: either argue against standard theories of reference, or argue that for 'moral responsibility' reference functions in an idiosyncratic way. Either horn of the dilemma

---

16 Blackburn (1993) has made a similar argument against error theories in metaethics. He argues that claims that there is widespread error in reference should ordinarily lead us to think that the theorist has simply misunderstood the core feature of moral language. Neither Hurley nor Blackburn say much about how we should accommodate fictional objects in these pictures of reference, nor do they seriously consider sophisticated forms of fictionalism about morality in general or responsibility in particular.

requires substantial argumentation, none of which seems to have been taken up by proponents of Strawson-like "impossibilist" views of moral responsibility and free will. In light of this, Hurley thinks that if we find ourselves taking seriously impossibilism, we should accept revisionism.

Hurley's account is framed as a conditional one: *if* one finds essentialist eliminativism plausible, then reconsideration of theories of reference should lead one to accept revisionism and not eliminativism. If successful, her arguments block those forms of eliminativism that appeal to impossible essences. However, they do not address all arguments for eliminativism. For example, Pereboom's (2001) hard incompatibilism does not claim that free will and moral responsibility are impossible. Instead, he argues that we should think it is overwhelmingly unlikely that we are free and responsible.

Despite the intended limits of Hurley's argument, the basic resources can be extended to put pressure on eliminativist accounts in general. Indeed, I take them to support the plausibility of a more categorical endorsement of revisionism about free will and moral responsibility.

Like Hurley, my development of a revisionist account has been motivated by considerations rooted in the philosophy of language, the fact of disagreements about intuitions, and the normative character of the concept of free will (Vargas 2001; 2005a; forthcoming*a*; Fischer et al. 2007). Perhaps more so than Hurley, I am inclined to think many of us do have incompatibilist intuitions in a variety of contexts, but that there is good reason to doubt that we are agents of the sort described by the best libertarian accounts. If the integrity of our normative practices rested on this form of agency, it would leave us in the morally precarious situation of blaming and punishing people on the basis of a picture of agency that we have little reason to accept—apart from our fervent hope

that we have it (Vargas 2009).

Fortunately, I believe there is an independent basis for various free will-characteristic practices, attitudes, and judgments, apart from libertarian agency. In light of its role in cultivating a special form of agency, I argue that there is adequate warrant for understanding free will in terms of various normative capacities. What makes the account revisionist is that I do not think this warrant is what we ordinarily believe licenses our responsibility ascriptions. Among other things, I reject the robust alternative possibilities requirement I believe to be embedded in some of our attributions of free will (Vargas 2008; forthcoming*b*).

On my account, free will is, roughly, the distinctive capacity of agents in virtue of which moralized praise and blame make sense. I argue that this capacity is best rendered in terms of the ability to recognize and respond to moral reasons. It is a picture on which the metaphysics of free will is determined not on the basis of primarily testing isolated pre-philosophical intuitions about free will, but instead by ascertaining what capacities would warrant specific normative practices, regardless of powers we tend to pre-philosophically associate with our agency (forthcoming*a*; forthcoming*b*). So, where Hurley's view has nothing to say to the possible-but-unlikely-to-have-free-will eliminativist, the version of revisionism I have offered is intended to provide a principled reply to all stripes of free will eliminativism. It does without appeal to the (I claim) dubious metaphysics of libertarianism or the conventional compatibilist's implausible insistence that compatibilist accounts are not at odds with important aspects of our self-conception.

I hold that the metaphysics of agency implicated in worries about free will are properly settled not by our intuitions but by our various practical, normative interests. In this, I maintain that the metaphysics we should look for is guided by fundamentally social and normative considerations. I am agnostic about how reference

sorts out: if our naive libertarian elements prove to be reference-fixing, I am a denotational revisionist compatibilist.[17] If, as I suspect, our commonsense libertarian commitments are not reference-fixing, then I am a connotational revisionist compatibilist.

Recent experimental work on intuitions about free will and moral responsibility have played some role in my thinking about these issues, leading me to think that ordinary thinking about free will may be fragmented in important ways (THIS VOLUME). The majority of non-philosophers are, I suspect, subject to distinct intuitions in different cases. That is, people feel the pull of incompatibilist intuitions more strongly in some cases. In other cases, they more strongly feel the pull of compatibilist intuitions.[18] None of this entails that there are no non-philosophers with

---

17 Does this make reference change too easy, ruling out the very possibility of eliminativism? I do not think so. Consider what we can say about the Christian God, if we were convinced of atheism. If God does not exist, presumably this is because is no candidate for denotational revision even if the concept serves some practical purpose precisely because there is no thing or property that plays even most of the main roles associated with God (e.g., creator of the universe *and* omnibenevolent agent *and* omnipotent *and* perfect *and* . . .). In contrast, with free will there are several candidate properties (think: what compatibilists point to) that do the primary work of the concept (e.g., licensing praise and blame, distinguishing between free and unfree action, warranting the reactive attitudes, etc.). Practical interests alone are not always sufficient for re-anchoring reference, although they are in this case because free will has a fundamentally practical dimension to it that God and, say, phlogiston do not. (N.B.: one *could* be a thoroughgoing instrumentalist or pragmatist about all the relevant concepts, but I do not mean to commit myself to such a view.)

18 Part of this may also be bound up in the responses to threats quite apart from determinism. Reductionism and the impugning of our folk psychological explanation likely play some role in driving reactions to cases that philosophers tend to read in terms of deterministic effects.

consistently incompatibilist intuitions, or that no one has consistently compatibilist intuitions. Still, I suspect that such persons constitute a minority. As in many aspects of human thinking, thoroughgoing consistency is the exception, not the rule.

(Elsewhere, I have suggested that this fragmentation may have ramifications for the development of professional philosophers' own views on free will (see Vargas 2006). Presumably, any given undergraduate comes to his or her philosophy class with mixed intuitions that, over time, get shaped by various internal and external pressures until they tend to assume a more consistent shape, whether compatibilist or incompatibilist. By the time that undergraduate becomes a Ph.D.-holding professional philosopher invested in the free will debate, the years of rational and disciplinary pressure in favor of consistency may have led that person to reify those intuitions in one direction or the other and to correspondingly dismiss, suppress, and oftentimes (even honestly) deny the presence of intuitions to the contrary.)

The idea that ordinary convictions contain both compatibilist and incompatibilist elements received worthwhile attention in a third self-described revisionist proposal. In a thoughtful and underappreciated article, Ira Singer (2002) argues that the intractability of free will debates can be explained by the presence of both genuinely compatibilist and genuinely incompatibilist instincts. Singer goes on to argue that we should acknowledge the fact of our libertarian convictions, and that we should excise them as best we can. This would leave us with a compatibilist-friendly concept and the attendant practices and attitudes. Notice that unlike Hurley, Singer is prepared to grant that the folk concept of freedom is incoherent. Indeed, he thinks that this is crucial for explaining the persistence of the debate. Moreover, he thinks the revisionist's conceptual surgery will have some

ramifications for our emotional life.[19] Although he is cautious about predicting just what emotional transformations follow from his account, he suggests that the reactive attitudes will be tempered in subtle ways by knowledge of the conceptual revision.

As Singer sees it, what prevents his proposed conceptual revision from being completely arbitrary is that it is yoked to what needs we have for distinguishing between freedom and our current historical and social circumstances. Consequently, he does not take it that we can straightforwardly repurpose any conventional compatibilist theory we like. Rather, *"within* compatibilism also there is a need to revise, to decide, to construct; we need to think about what compatibilist views best serve our various needs and purposes" (38).[20] Conventional compatibilist accounts may provide a starting point, but they cannot be ending points.

Besides Singer, Hurley, and myself, there are a variety of views in the current literature that are plausibly revisionist in conception or implication. For example, Henrik Walter (2004) has explicitly embraced a revisionist position on free will and moral responsibility in light of developments in neuroscience. Nichols (2007a; 2007b, 2008) has argued that even if our ordinary (libertarian) conception of agency is untenable, many of the attitudes and practices that were supposed to depend on that conception of agency are largely insulated from changes in our metaphysical beliefs. Clarke (2010), a former proponent of agent-

---

19 Where Hurley and I are largely content with a kind of conceptual revisionism, Singer embraces "emotional revision" (40). I suspect that there might be some modifications at the margins of our ordinary moral life, but I am dubious about changes on the scale Singer seems to think will follow.

20 Partly influenced by Singer, I have noted a distinct but related concern about conventional compatibilist accounts under the guise of "revisionism on the cheap" (Fischer et al. 2007, pp. 152-153).

causal libertarian, has recently suggested that aspects of our self-conception, especially as they relate to moral responsibility "might need alteration in some unacknowledged respects."[21] This suggests the possibility of more diverse forms of revisionism than I have gestured at here. Where one might not embrace revisionism about free will, one might embrace it with respect to some other issue connected to the philosophical literature on free will.[22]

5. Ramifications and challenges

Before concluding, I wish to briefly mention some ongoing challenges and unresolved issues facing revisionist accounts of free will.

First, there is the matter of the correct diagnostic account of our pre-philosophical commitments on freedom and moral responsibility. A revisionist account requires some conflict with our received free will commitments. It therefore matters what ordinary or folk beliefs about free will come to. Most current revisionist accounts have been revisionist compatibilisms, motivated by the convinction that folk beliefs contain incompatibilist elements. It would be problematic if it turned out that ordinary persons did not have incompatibilist commitments.

In recent years there has been a growing body of experimental work designed to tease out ordinary convictions about free will and moral responsibility (THIS VOLUME). One notable

---

21 See, also, Clarke (2009): "it's a harder problem to see how genuinely free and responsible action is possible. I do think some revision of our everyday understanding of our responsibility for what we do may be required" (26).

22 Alternately, one might embrace revisionism about free will on grounds distinct from threats via determinism. Nahmias may accept a species of revisionist compatibilism about free will and *reductionism*, while regarding himself a conventional compatibilist about free will and determinism.

result of that work has been the discovery that ordinary persons will in many cases have strongly compatibilist reactions to a wide range of prompts. The research into these issues is ongoing, but it raises complicated questions about the extent to which self-described revisionist compatibilists are genuinely revisionist if ordinary persons are mostly or largely compatibilist in their commitments. So, unless one accepts something like Singer's claim that we already need to re-write our compatibilist accounts in light of a clearer connection to our practical and normative purposes, empirical results that overwhelmingly favored a compatibilist diagnosis would be a serious challenge to at least the revisionist element of these accounts.

Still this concern may prove to be only superficial. For any self-described revisionist account that proves to not be revisionist, we would still have a substantive prescriptive account that merits consideration on its own terms. If the best self-described revisionist compatibilist account turns out to be nothing more than an excellent conventional compatibilist account, this would, I suspect, hardly dismay the account's proponent.

A second challenge to revisionists is to articulate the grounds on which revision is licensed. Without some account of what it is that anchors the revision, revisionism threatens to collapse into an "anything goes" account on which the rules guiding revision are impossible to make out and the products of revisionist theorizing are correspondingly impossible to evaluate. It is well and good to appeal to the general idea of our practical and normative interests, but articulating what those interests are and how they structure a revisionist proposal is difficult work that few have attempted to undertake. This general approach also raises questions about the relationship of theorizing about free will with commitments in normative ethics and metaethics (McKenna 2009; Double 1996). If an account of free will is beholden to practical and normative

interests, we might wonder about the extent to which one can adequately develop such an account in isolation from robust commitments in these other areas of philosophy. Grappling with these questions remains a significant burden for any revisionist account.

A third, and perhaps related challenge is to articulate the extent to which a given revisionist proposal entails revisions in our practices and attitudes. These issues are difficult, hinging on the extent to which particular practices depend on the jettisoned conceptual elements. As noted above, Singer suggests a picture on which conceptual revision might help to attenuate the reactive attitudes in the same way Hume thought the theoretical conviction that stoicism is true might weaken the force of some desires. However, there are those have taken a stance of general skepticism about transformation of practices in light of metaphysical commitments. Mark Johnston (1992) has argued that "although ordinary practitioners may naturally be led to adopt metaphysical pictures as a result of their practices, and perhaps a little philosophical prompting, the practices are typically not dependent on the truth of the pictures. Practices that endure and spread are typically justifiable in non-metaphysical terms" (590). He goes on: "That the practice of attributing responsibility depends for its justification on facts about free agency . . . does not settle it that the practice of attributing responsibility depends for its justification on facts about uncaused causings. The picture of free willings as uncaused causing may have only a minimal role. It may yet be epiphenomenal to the practice" (592). Johnston's pictures is one on which our practices may be completely insulated from our metaphysics.

There is ample room for positions between Singer and Johnston's views. For example, one could hold that there ought to be changes at the margins of our practices, thus leaving the bulk of

our practices and attitudes unchanged. In general, if we identify elements of our practices that depend on conceptual content that has been revised away, we have a candidate for revision in our practices. Here, though, we would do well to remember that even if some aspect of a practice had its *origin* in a conviction we now abandon, it does not follow that the function or current meaning of the practice depends on that conviction. We might find that some elements that appear to presume the truth of a to-be-excised conviction can be justified on independent grounds.[23] Indeed, this seems to be the central strategy for at least some strands of revisionist approaches to free will. How this goes, though, is the nub of it.

It may be comparatively easy to motivate a revisionist account of responsibility (e.g., by appealing to some metaphysically innocuous justification for praise, blame, and desert attribution). However, if one construes free will in terms that disconnect it from its role in moral responsibility, then it is less clear what it is that would provide independent grounds for holding on to those practices, attitudes, or inferences that depend on the too-be-jettisoned conceptual elements of our folk notion of free will. In turn, this raises the specter that revisionist accounts of free will that countenances significant changes in our practices and attitudes are not so different from eliminativist accounts, after all. Thus, revisionists who accept significant changes to our practices and attitudes will need to be careful to say how and why these accounts

---

23 Shaun Nichols (2007a) has argued that there are good reasons for agents to have and express retributive impulses, and that these reasons hold independent of whether we are libertarian agents. If so, then revision away from libertarianism would not necessarily entail a loss of retributive attitudes, even if one thought that as a matter of ordinary course we maintain that only libertarian agency would suffice to support retribution.

are different than more familiar eliminativist accounts.

I have mentioned three challenges facing revisionist accounts: disputes about the diagnosis, the grounds for revision, and what conceptual revision entails for practices and attitudes. There are obviously other challenges facing revisionists. Some critics will surely object that a given revisionist view presumes a false picture about the semantics of free will, or that revisionist accounts amount to an illegitimate changing of the subject, even if such wretchedness is no subterfuge. Others will dispute the idea that we can make sense of the idea of "the work of the concept" in a way that permits conceptual revision without conceptual abandonment. Moreover, I have said little about how revisionism about free will connects to revisionism about moral responsibility, desert, and retribution. More work is in order.

My sense is that revisionist views will continue to proliferate, at least for a while. In a field as intricately developed as free will, revisionism offers something unusual: relatively undeveloped territory with plenty of low-hanging fruit. I also suspect that the serious regard in with which eliminativist views have only recently come to be viewed provides a natural segue for reflection on revisionism. Anecdotally, philosophers tend to come to revisionism only after having first espoused eliminativism. Regardless of how philosophers come to it, revisionism will surely benefit from continued philosophical labor.[24]

---

24 Thanks to the following folks for helpful feedback on this essay or ideas in it: John Fischer, Robert Kane, Jason Miller, Shaun Nichols, John Perry, Dan Speak, and Kevin Timpe.

References

Bennett, Jonathan. 1980. *Accountability*. In *Philosophical Subjects*, edited by Zak Van Straaten. New York: Clarendon.

Blackburn, Simon. 1993. *Errors and the Phenomenology of Value*. In *Essays in Quasi-Realism*, Oxford: Oxford University Press.

Clarke, Randolph. 2009. *Randolph Clarke*. In *Philosophy of Action: 5 Questions*, edited by Jesús Aguilar, and Andrei Buckareff. USA and UK: Automatic Press/VIP.

Clarke, Randolph. 2010. Determinism and Our Self-Conception. *Philosophy and Phenomenological Research* 80 (1): 242-50.

Dennett, Daniel. 2003. *Freedom Evolves*. New York: Viking.

Double, Richard. 1996. *Metaphilosophy and Free Will*. New York: Oxford University Press.

Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas. 2007. *Four Views on Free Will*. Malden, MA: Blackwell.

Heller, Mark. 1996. The Mad Scientist Meets the Robot Cats: Compatibilism, Kinds, and Counterexamples. *Philosophy and Phenomenological Research* 56 333-37.

Hurley, S.L. 2003. *Justice, Luck, and Knowledge*. Cambridge, MA: Harvard University Press.

Hurley, Susan. 2000. Is Responsibility Essentially Impossible? *Philosophical Studies* 99 229-68.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York: Oxford University Press.

Johnston, Mark. 1992. Reasons and Reductionism. *The Philosophical Review* 10 (3): 589-618.

Lewis, David. 1973. *Counterfactuals*. Cambridge: Harvard University Press.

McKenna, Michael. 2009. Compatibilism and Desert: Critical Comments on *Four Views on Free Will*. *Philosophical Studies* 144 3-13.

Mele, Alfred R. 2009. *Effective Intentions: The Power of the Conscious Will*. New York: Oxford University Press.

Nichols, Shaun. 2006. Folk Intuitions on Free Will. *Journal of Cognition and Culture* 6 (1 & 2): 57-86.

Nichols, Shaun. 2007a. After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes. *Philosophical Perspectives* 21 405-28.

Nichols, Shaun. 2007b. The Rise of Compatibilism: A Case Study in

Quantitative History of Philosophy. *Midwest Studies in Philosophy* 31 260-70.

Nichols, Shaun. 2008. *How Can Psychology Contribute to the Free Will Debate?* In *Are We Free? Psychology and Free Will*, edited by John Baer, James C. Kaufman, and Roy F. Baumeister. New York: Oxford University Press.

Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge.

Pereboom, Derk. 2005. Defending Hard Incompatibilism. *Midwest Studies in Philosophy* 29 (1): 228-47.

Pereboom, Derk. 2009. Hard Incompatibilism and Its Rivals. *Philosophical Studies* 144 21-33.

Scanlon, T. M. 1988. *The Significance of Choice*. In *The Tanner Lectures on Human Values*, edited by Sterling M. McMurrin. Cambridge, UK: Cambridge University Press.

Singer, Ira. 2002. Freedom and Revision. *Southwest Philosophy Review* 18 (2): 25-44.

Smart, J.J.C. 1961. Free Will, Praise, and Blame. *Mind* 70 291-306.

Strawson, Galen. 1994. The Impossibility of Moral Responsibility. *Philosophical Studies* 75 5-24.

Strawson, P. F. 1992. *Analysis and Metaphysics*. New York: Oxford.

Strawson, P.F. 2003. *Freedom and Resentment*. In *Free Will*, edited by Gary Watson. New York: Oxford University Press.

Vargas, Manuel. "Even Better Than the Real Thing: Revisionism and Responsibility." Ph.D. Dissertation, Philosophy, Stanford University, Stanford, CA.

Vargas, Manuel. 2004. Responsibility and the Aims of Theory: Strawson and Revisionism. *Pacific Philosophical Quarterly* 85 (2): 218-41.

Vargas, Manuel. 2005a. The Revisionist's Guide to Responsibility. *Philosophical Studies* 125 (3): 399-429.

Vargas, Manuel. 2005b. Compatibilism Evolves? On Some Varieties of Dennett Worth Wanting. *Metaphilosophy* 36 (4): 460-75.

Vargas, Manuel. 2006. Philosophy and the Folk: On Some Implications of Experimental Work for Philosophical Debates on Free Will. *Journal of Cognition and Culture* 6 (1 & 2): 239-54.

Vargas, Manuel. 2008. *Moral Influence, Moral Responsibility*. In *Essays on Free Will and Moral Responsibility*, edited by Nick Trakakis, and Daniel Cohen. Newcastle, UK: Cambridge Scholars Press.

Vargas, Manuel. 2009. Revisionism About Free Will: A Statement & Defense. *Philosophical Studies* 144.1 45-62.

Vargas, Manuel. Forthcoming a. Responsibility in a World of Causes. *Philosophic*

Exchange

Vargas, Manuel. Forthcoming b. *Situationism and Moral Responsibility: Free Will in Fragments*. In *Decomposing the Will*, edited by Till Vierkant, Julian Kiverstein, and Andy Clark. New York: Oxford University Press.

Vargas, Manuel. Forthcoming c. *The Revisionist Turn: Reflection on the Recent History of Work on Free Will*. In *New Waves in the Philosophy of Action*, edited by Jesus Aguilar, Andrei Buckareff, and Keith Frankish. Palgrave Macmillan.

Vargas, Manuel. 2004. Libertarianism and Skepticism About Free Will: Some Arguments Against Both. *Philosophical Topics* 32 (1&2): 403-26.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Walter, Henrik. 2004. Neurophilosophy of Moral Responsibility: The Case for Revisionist Compatibilism. *Philosophical Topics* 32 (1&2): 477-532.

White, Stephen L. 1991. *The Unity of Self*. Cambridge, MA: MIT Press.