HOW TO SOLVE THE PROBLEM OF FREE WILL

Forthcoming in Russell and Deery, eds. (2013) *The Philosophy of Free Will: Essential Readings From the Contemporary Debate*, OUP.

Manuel Vargas, University of San Francisco
mrvargas@usfca.edu
version: 11.25.12

I. Take-backs

Let me start off by acknowledging that the title of this essay is both presumptuous and completely misleading.

First, I don't think there is anything that has sole claim to being *the* problem of free will. There are a number of distinct philosophical puzzles that have gone under the label "the problem of free will." For example, there is the question of what sort of power or variety of control we need in order to be properly held morally responsible. There is also the question of what sort of freedom we must have for us to have true beliefs about our powers when we deliberate about what to do. There is also the question of whether and how it makes sense to say that we cause things, or initiate new chains of events, as opposed to our just transmitting the effects of prior causes. You *could* think that the answer to all these things—and others I haven't mentioned—comes to the same thing. However, I am inclined to think that they do not, and that anyway, we ought not start off by supposing that they do.

Second, I'm not really going to try to *solve* any of the problems of free will. Instead, I am going to show *how* we can solve it. The idea is to give you a general recipe or formula for constructing a solution to it. Elsewhere, I have tried to bake that cake given by the recipe, but my aspiration here is much more limited. I just want to convince you that this recipe looks good enough to be worth trying out on your own time.

Third, I'm aware that all of this sounds *really* presumptuous. Lots of thoughtful folks have thought really hard about the various problems of free will. Many of them have offered purported solutions to the prob-

lem of free will. However, it isn't as though any of those accounts has widespread support. So, you might wonder, what makes this account so special?

In reply: well, yeah, this looks bad. Nevertheless, I think we've been very close to an adequate solution to at least one version of the free will problem for some time now, and all we have needed is a distinction or two and a bit more clarity about some methodological issues. My contribution to these matters is to put a few pieces in place that allow us to see our way through to a solution, the larger pieces of which have been developed by many others.

So now that I've taken back nearly everything promised in the title, I should say what I'm actually going to do. First, I'll say a bit about conventional solutions to one version of the free will problem. My focus will be on what is dissatisfying about them. Then, I'll consider whether we should just accept that we don't have free will. I'll argue that we shouldn't, because there is a possibility—and a plausible one at that—that we have overlooked. I'll then discuss the shape of this possibility and how it provides us with the outlines of a satisfactory resolution to the free will problem.

2. The problem with which we are concerned

The version of the problem of free will I am interested in trying to solve is the problem of explaining what free will is and whether we have it, where by free will I mean a feature, power, or ability of responsible agents that is especially distinctive of them being morally responsible.

I say "especially distinctive" rather than "required" because there are many features of agents—beliefs, desires, intentions, emotions, social connections, consciousness, and so on—the absence of which may preclude moral responsibility in some or another circumstance, but that are nevertheless not distinctive of specifically responsible agency. However, for my purposes, free will is the kind of thing that shows up in distinctively (morally) responsible agents. In particular, something about free will should help explain why agents with it can be properly held responsible.

There are some things I could say to motivate such presumptions, but I'm content to leave them as undefended premises in what follows.

In focusing on a responsibility-centric conception of free will, it behooves me to clarify that the variety of responsibility that is at stake here is *moral* responsibility. I am interested in a notion of responsibility connected to the licensing of moral praise and blame—blameworthiness, you might say. This is distinct from, for example, causal and legal responsibility. I might be worthy of moral blame for snickering loudly at your choice of trousers or your manner of speaking, but it would not follow that under the law there is anything for which I am thereby responsible. Causal and moral responsibility come apart as well: a bad alternator might be causally responsible for my car not starting, but there is no moral fault to be found with the alternator.

One complexity about moral responsibility is that it admits of both direct and indirect cases. When I surreptitiously put beef stock in food for vegetarian friends, I am directly responsible for serving them non-vegetarian food. But there are outcomes of my choices for which I can be responsible that are less certain but morally salient possible outcomes of what I do that I foresee, or could foresee. So, if I plan to pursue a high speed road race immediately following an extended tequila-shooting contest with my drinking team, it seems plausible to think that I will be responsible for a wide range of disastrous results that will follow in the wake of my undertaking that two stage adventure. When I pass out at the wheel, moments before my car goes shooting off a cliff into the Pacific Ocean, I will not be absolved from responsibility when my teammates point out that I was unconscious at the time I went off the cliff. If still sufficiently articulate, they might belligerently proclaim that consciousness is widely regarded as a requirement for responsibility, and I was manifestly unconscious when my car left the road. But we would do well to ignore them. After all, my responsibility here is very real, if indirect.

So, in the following, I am interested in a distinctive agential power, a power that both directly and indirectly supports moral responsibility.

3. The standard solutions and what's wrong with them

There are a number of purported solutions to "the" problem of free will, but two in particular have garnered the lion's share of adherents: *libertarianism* and *compatibilism*.

Libertarianism is ordinarily characterized as the conjunction of two different commitments. The first is *incompatibilism*, or the view that free will is incompatible with determinism.[1] The second is a commitment to the *existence* of free will. So, a libertarian thinks that if determinism were true, we wouldn't have free will. But we do have free will, insists the libertarian, which commits him or her to the view that we are not determined.

Although historically significant, and not without conceptual utility, I think the focus on determinism is more than a little misleading. There aren't many physicists or philosophers who think that determinism is true as a description of low-level physical particles. (Oddly, social scientists remain great enthusiasts of determinism.[2]) So it should strike you as a bit strange to worry about whether free will is incompatible with determinism.

Instead, it seems to me, the real concern should be whether we possess forms of agency sufficient to sustain practices of moralized praising and blaming, given a scientifically-informed, independently plausible picture of the world. After all, even if determinism is true, there might be any number of other things that could threaten our having free will—whether it is neuroscience, discoveries about how the psychology of our

1. Definitions of determinism are plentiful. For our purposes, we can assume that determinism is the view that (post-Big Bang) everything has a cause, and that causes are not probabilistic in any interesting way (i.e., the probability of a given cause bringing about an effect is always either 1 or 0). So, a post-Big Bang world that had uncaused events would not be deterministic. Nor would a world in which there were sometimes causes that had, say, only a 75% probability of bringing about some effect.

2. Chris Franklin has persuasively argued that much of the grip that determinism has exercised over the imaginations of scientists has arisen out of a concern to expunge uncertainty from their models and out of a confused (if historically influential) model of science (Franklin Forthcoming).

agency functions, and so on. Determinism is only one threat, and perhaps the least scientifically credible threat, at that.

Still, determinism has a kind of pride of place in these discussions. One reason is that it remains useful as a way of characterizing an influential class of views. Even if many of us remain suspicious about determinism as a characterization of the universe as a whole, it remains possible that we might learn that determinism, or something very close to it, obtains in the parts of the universe in which our agency functions. What the libertarian insists is that were we to learn that our actions were determined in this way, that would be sufficient to show that we are not free and responsible.

Compatibilism is the other view with a plausible claim on a plurality of adherents, at least amongst those who have joined these debates. By compatibilism, I mean the view that free will is compatible with determinism. In recent years, there has been some dispute about the scope of this and related terms, for example, whether it is possible that one can be a compatibilist about moral responsibility without being a compatibilist about free will. Although there is a good deal to say about these matters, I will bracket them. For present purposes, I will use 'compatibilism' and 'incompatibilism' to refer to the view that *both* free will and moral responsibility are, respectively, compatible and incompatible with determinism.

I claimed that in this section I would say why the standard solutions to "the" problem of free will don't work. Nothing I will say here is immune to reply by even an average defender of the views I criticize. Still, my hope is to make some remarks that will resonate with those who have cast their eyes at least once on the standard approaches, and walked away not entirely convinced.

Libertarianism does a fine job of capturing many threads of ordinary, commonsense thinking about the nature of our agency in the world. There is a growing body of experimental research that shows that when people reflect on questions of free will in the abstract, many of us tend to have broadly incompatibilist reactions (Nichols and Knobe 2007; Roskies and Nichols 2008; Sarkissian et al. 2010). And, overwhelmingly, people seem to think that even if the rest of the world is deterministic, at least human choice-making is not (Nichols and Knobe 2007). So even if not

everyone is pre-philosophically (or perhaps more accurately, at the first stages of doing philosophy) a libertarian when thinking about matters in the abstract, a good many people seem to be.

Libertarianism has a second virtue. It piggy-backs on the considerable (if imperfect) virtues of classic arguments for incompatibilism. I won't rehearse them here, but there is a body of powerful arguments for incompatibilism. They can be resisted in various ways, with various degrees of dialectical sophistication. However, incompatibilism gets some claim to being intuitive from the fact that familiar arguments for incompatibilism do seem to codify a very natural way of thinking about free will. What it takes to resist these arguments is oftentimes a very subtle interpretation of powers, laws of nature, abilities, and so on. The nuances of these views, and the complex arguments it takes to make such views credible to even other philosophers, seem to me to be prima facie considerations against their credibility as descriptions of ordinary convictions.

Finally, I think the intuitiveness of libertarianism is perhaps connected to a diverse set of other influences. The legacy of dualism, various religious traditions, and a tendency to read off one's metaphysics from confusions about phenomenology and the explanatory frameworks for human action have all given libertarianism a foothold in our imagination of human agency.

Despite its considerable intuitive appeal, libertarianism has several shortcomings. I will mention two: doubts about its plausibility and worries about the moral cost it carries.

First: once you see what it takes to make good on the commitments of libertarian pictures of agency—the postulation of a special and radically different metaphysics for humans (e.g., agent causal libertarianism), the existence of indeterministic mechanisms in very specific places in human deliberation and not others (event causal libertarianism), or the postulation of uncaused events in the production of human action (uncaused event libertarianism)—it is easy to feel uneasy. The accounts can feel like armchair neuroscience, or they seem to invoke an entirely *ad hoc* metaphysics, postulated solely to preserve a picture that we have no independent motivation for believing in, beyond the fact that we seem greatly invested in perpetuating practices of moralized praise and blame. And

that, I think, is a real problem. If we were to sit back and ask ourselves what we have good reason to believe about our agency, solely on the evidence, it is hard to imagine that we'd come up with libertarianism.

Now this isn't really an argument against libertarianism. For all I've said, libertarians might be right. Our agency might indeed have whatever properties are described by your favorite libertarian theory. Still, I've never seen good evidence for thinking that we are, in fact, agents of the sort libertarians describe. At best, we get hints that, for example, some brain processes are not deterministic. But this feels a bit like trying to prove the existence of fairies by noting that there are some small creatures with translucent wings. In neither case is consistency with the known facts an argument for the actuality of the additional posit. And there is almost always an additional posit; no libertarian thinks that raw indeterminism is sufficient for free will. The indeterminism has to happen in particular ways at particular times, and this is precisely what there is virtually no evidence for. So, it seems, libertarianism is an under-motivated attempt to preserve a pre-scientific picture of humans as radically distinct and separate from the physical, material world.

There is a second reason to worry about libertarianism. The worry is this: given the fact that libertarianism has little or no evidence in its favor, beyond our hope that it is true, adherence to libertarianism seems to undermine our grounds for holding one another responsible. That is, it makes the justification for our blaming and punishing hinge on the hope or aspiration that we are libertarian agents, a hope or aspiration for which *there is no positive evidence and considerable disagreement about whether anyone possesses the requisite power* (Pereboom 2001, pp. 161, 198-199; Pereboom 2006, pp. 562-64; Vargas 2009).

A concrete example may make this point clearer. Consider *Fiery*. Fiery is a skeptical subject of a significant moral blame, and likely, punishment. Perhaps she faces the death penalty, if that is permissible, and if not, then some very significant censure where that variety or some large quantum of that censure (whether blame or punishment) depends on the presumption of her being a libertarian agent. Now, let us imagine that Fiery demands to know *why* such treatment is justified.

Her libertarian persecutor must acknowledge that we have no evidence to support the hope that underwrites our treatment of her—that is, the hope that Fiery is, indeed, a libertarian agent. But Fiery will surely protest: the mere *possibility* that she deserves some extra quantum of blame or punishment beyond that required for say, rehabilitation, does not, by itself, make such treatment justified. After all, Fiery insists, there is also a chance that she—and everyone else—might not be libertarian agents. Indeed, this strikes her (especially now!) as considerably more plausible than her prosecutor's insistence that libertarianism is true.

Fiery is surely right about this much: if she is indeed not responsible, it would be grossly unjust to hold her accountable to any degree beyond the degree of blame and punishment warranted by non-libertarian considerations. On the presumption that one should avoid gross injustice when one can do so, and that it is wrong to blame when there is no evidence that the target is responsible, the only defensible course of action is to abandon holding Fiery (and everyone else) responsible in whatever degree the presumption of libertarian agency entails. So, it seems to me, we had better have a justification for blame that runs deeper than the wish or hope that we are libertarian agents.

So: libertarianism's purported solution to this version of the free will problem is not much of a solution.

Let us turn to consider the other purported solution to the problem of free will: compatibilism. Like libertarianism, compatibilism has some claim of capturing our pre-philosophical convictions. For example, there is a body of research that shows that when it comes down to brass tacks, when we think about concrete, particular cases in which there is a clear victim, we seem quick to blame and condemn regardless of whether or not we are told the world is deterministic (Nahmias et al. 2006; Nichols and Knobe 2007; Woolfolk et al. 2006).

So, it seems, compatibilism has some claim on us too. The dialectic here is complicated, and more complicated than I can adequately address in this essay. Here's the upshot, though: for all of compatibilism's virtues, it will always seem like a cheat to a good many of us. Many of us persist in having incompatibilist convictions. Although there are clearly cases in which a strong majority of folks give compatibilist reactions to

various prompts, incompatibilist reactions never completely disappear, remaining at roughly a quarter to a third of responses among the populations where these things have been studied.

We could argue about which set of cases is more revealing—perhaps there is an error in our thinking caused by emotion-triggering cases with personalized victims, or perhaps it is only when emotions are engaged that we see the true moral significance of an act—but I think there is no easy path to the resolution of this dispute. Instead, I think we should just accept that (for good or bad reasons) as a matter of ordinary convictions we have a mixed picture of the requirements for moral responsibility. If that is right, though, then any philosophical account of moral responsibility that is going to satisfy those convictions will need to sometimes invoke libertarian convictions.

Some compatibilists have insisted that one of the appealing features of compatibilism is that it does not make our moral responsibility "hang on a thread" (Fischer 2006, p. 6). And, one might think, that such considerations might give us some reason for jettisoning lingering libertarian convictions. But we must be careful here. It might be appealing if our theory of free will ensured that we are responsible, but this doesn't (by itself) seem like a good reason to think that a thoroughgoing compatibilist theory is *true*. After all, I find it appealing to think that there is an afterlife awaiting me, perhaps with an endless supply of the finest mezcals, bourbons, and Belgian beer (plus, my spouse and kids . . . and friends and extended family (that I like) and video games, quite plausibly). But as widespread a view about the afterlife as this might be (allowing for substitution about the precise details), we might wonder if the fact of its being appealing and its possibility of playing an organizing role in my life should yet constitute a reason for thinking that such an afterlife exists.

So, I think conventional compatibilist theories are in an odd position. They are plausible only to the extent to which ignore the very convictions that gave us a problem that needs solving. The way they solve the problem is by denying we ever had the problematic commitments in the first place. And, independent appeals to how nice it would be to insulate our moral practices from threat does not seem to constitute a very compelling reason to think the view is true.

To sum up: the problem with standard solutions is that they don't seem to do a very good job of solving that problem. Libertarians offer us accounts that would be nice to believe, but they don't give us very good reasons for believing in them. Compatibilists offer us solutions, but solutions that seem to work only by insisting there was no real problem to be solved in the first place.

Given that conventional solutions seem unfortunately aspirational or disingenuously evasive, we might conclude that the only solution is to acknowledge that we do not have free will. Indeed, one might think, far too many philosophers have been callow about this, scared of embracing the tough but radical conclusion. So, we might think, at least *we* shall be tough enough to stare the problem in its face and deal with it honestly and with integrity.[3]

Perhaps there is a matter of disposition here: some philosophers are drawn to more radical conclusions and others regard such conclusions as proof of an argument gone wrong. My own view is that we can be dissatisfied with conventional libertarian and compatibilists accounts, and still think that the no free will view is woefully undermotivated. Let me explain.

4. What's wrong with the nay-sayers

We can make some progress by starting with a distinction between *diagnostic* and *prescriptive* theorizing.

When we face a philosophical puzzle, we can try to provide a diagnosis of what is going on. In providing that diagnosis, we offer a description of the state of our thinking, and ideally, an explanation of how we came to have the problem that exercises us. So, a diagnostic account of free will is one that endeavors to describe the contours of our thinking that have given rise to the problem of free will.

3. As Nietzsche once noted: "It is certainly not the least charm of a theory that it is refutable; it is precisely thereby that it attracts the more subtle minds. It seems that the hundred-times-refuted theory of the 'free will' owes its persistence to this charm alone; some one is always appearing who feels himself strong enough to refute it" (Nietzsche 1966, § 18).

However, we might pursue a different project. We could endeavor to provide an account of how, all things considered, we ought to think about some subject matter. This may or may not overlap with how we currently think about things. Sometimes how we ought to think about things just is how we think about them. Other times, what we conclude is that we ought to think about things somewhat differently than we do. So, for example, at one point in time a popular view was that water was one of the four basic indivisible substances of the universe. When the chemical theory of water gained acceptance, presumably these people did not conclude that water did not exist. Rather, they concluded that the nature of water was different than they had imagined it to be.

Don't let the example from the history of science mislead. Moral, social, and legal categories have all changed over time as a result of various pressures—some empirical, some conceptual, some normative. Few of our received notions have remained unmolested by the expansion of human learning and the accretions of diverse cultural practices.

It is the possibility of a gap between what we think and what we ought to think that offers us a way out of the familiar debates about free will. Here's my suggestion: we can resolve a number of problems familiar to us under the rubric of 'free will' if we permit ourselves to take seriously the possibility that free will might not be the sort of thing we supposed it to be. Moreover, once we have seen how such an account might go, and we reflect on our pre-philosophical convictions as we find them, what we should conclude is that we have no good reason to hold on to various presuppositions about free will, presuppositions that have precisely led to our conception of free will *as a problem*.

In suggesting this possibility, I don't mean to downplay the fact that such an account would leave us with some hard work. For example, we would need some story of how we can go about characterizing free will without just appealing to our received convictions or the intuitions we find ourselves having. (My answer: think about the role of the concept, the work it does, and what notions and practices it regulates.) And, we would need some reason to think that the proposed prescriptive account is an account of free will and not some other things. (Roughly, the test is to ask whether it is capable of doing the bulk of the conceptual and social prac-

tice-coordinating roles we associate with free will). But if we can provide reasonable enough answers to these things, we move tantalizingly close to a resolution of the exhausting and exhausted-seeming debates about free will.[4]

In general, there are at least two classes of cases where jettisoning significant convictions widely associated with some notion have been appealing: (1) in the case of scientific discoveries (e.g., the advent of the chemical theory of water displacing, say, broadly Empedoclean and Aristotelian pictures of water) and (2) in the case of ideas whose function or role in our life has to do with social regulation, and where background presumptions and/or motivating social pressures have shifted sometimes in response to empirical data and/or to new forms of social organization. Examples include, perhaps, marriage, social class position (e.g., as reflecting stratifications reflective of God's favor), race (as a biological essence vs. a social kind), and so on. In each of these cases, we came to fundamentally re-conceive the matter. Indeed, the whole of morality might have plausibly been subject to such transformation in conviction.

Recall the state of ethics among the European intelligencia at the end of the 19th century. God's obituary was being written, and more than a few thoughtful people, including Nietzsche and Dostoevsky, wondered whether God's non-existence would entail that "everything is permitted." Here's one reason people might have thought that: if you thought that the nature, content, and significance of morality is essentially settled by God's will, and if you thought God doesn't exist, then moral claims will likely look to you to be false, nonsense, or at least not binding.

There are different ways to read the subsequent history of ethical theorizing. On one reading, we are still working out the fact that without God, there is no hope for a adequate foundation for morality. A different, and perhaps more common view (at least in the overwhelmingly atheist-in-

4. Notice that if we accept the possibility of some notion's nature being other than we ordinarily conceive it to be, this does not preclude nihilism-warranting discoveries. If nothing or too many diverse things play the relevant realizer in the world, this would be reason for taking seriously the nihilist's recommendation. On this approach, eliminativism is not eliminated, but its likelihood is reduced.

clined world of Anglophone philosophy) is the thought that the proper lesson from the death of God was not that morality doesn't exist, but that the foundation of morality was rather different than many people supposed. Rather than God's existence and His commands being essential to morality, it turns out that something else (say, what we would universally will under full information in ideal conditions, or what would produce the greatest welfare, or what we would agree to if concerned to create a system of cooperation, or . . .) is the core of morality.

This basic pattern of concept change should seem, with a bit of reflection, somewhat familiar: we have a cherished way of thinking about something, a way that helps make sense of our lives and orders important practices; it becomes threatened by some new consideration or an overturning of some old presumption; people rush to declare that the old notion is now bankrupt and to be rejected; but then others rush in to advocate a transformation or rehabilitation of the old view, which now rejects the troublesome element or presupposition that was previously thought to be essential to the notion. Sometimes the rehabilitation sticks and sometimes it doesn't. Whether it does is a function of any number of factors, but at least in the case of ideas that play some role in social regulation, one prominent factor is surely whether or not the proposed transformation still permits a similar kind of social regulation as the pre-revised notion.

So now, finally, we get to the heart of the matter: can we go in for a revised notion of free will, one that does without the (to some, apparently essential) incompatibilist elements? That is, is our notion of free will the kind of thing that is enough of a social regulation-like idea to admit to this sort of transformation? I think so.

Recall an idea I mentioned at the beginning, that there are various notions of free will that philosophers have bandied about. I'm not sure that all the notions out there that have some claim on being labeled 'free will' are sufficiently social regulation-ish to make this idea plausible in all of those cases. However, if we focus on the notion of control that is required to license moral praise and blame, then the answer is very plausibly *yes*. This idea, that we are concerned with something intimately tied to social practices, gives those of us with *this* interest—i.e., free will understood

in terms of a condition on moral responsibility—some reason to think we should be *revisionists* rather than eliminativists or nihilists about free will. That is to say: we can think that a philosophically adequate account of free will will conflict with aspects of our ordinary commitments.[5] Crucially, the revisionists holds that such conflicts are permissible (and indeed, predictable) whenever we reject a particularly troublesome aspects of our ordinary views—say, libertarian commitments—,because we've found some other way to secure what is at stake.

Of course, there are some who will insist that, manifestly, free will should *not* be understood in this way. Perhaps you are among them, thinking that free will is a kind of property of our agency that we have or don't, and that it can be analyzed independently of moral or social concerns. If this is nothing more than noting that you are interested in a different notion of free will than the one with which I am concerned, that's fine. Perhaps I want to think about chocolate and you want to think about sauerkraut. But suppose you intend this as an objection to the same notion I'm interested in, and that this notion is to be understood in terms of some metaphysical property, the status of which is independent of anything in the realm of morality or social regulation. On your view, the permissibility of the moral concerns *presumes* or *depends* on whether or not we have this metaphysical property. Call this *the metaphysical reading*.

Here, there are two things to say. First, we can repeat the dialectic: we might *think* some independent metaphysical nature is the core of a responsibility-centric notion of free will, but we could be in error about this, and an otherwise plausible account of free will that dispenses with this presumption is partly an argument against the presumption of the metaphysical reading. Second, even if we grant that the metaphysical reading is an essential part of something we in fact care about, it remains open to us to object that this is merely another one of a long line of mysterious properties according to which we regulate our social and moral lives, but

5. More precisely: a theory is revisionist whenever its prescription *conflicts* with the diagnosis. Cases in which a theory merely refines some commitment or the theory stakes a claim or injunction on some subject about which we have no antecedent commitments is not revisionist in the sense with which I am concerned.

that turns out to be otiose to the concerns for which we've postulated the property. In this, the metaphysical conception of free will might be like blood purity, succubi, immaterial souls, and the divine right of kings: you *could* believe in such things, but we don't *need* to believe in such things if, for example, we wish to regulate our social lives in ways that are both mutually justifiable and that permit us to flourish.

So, *if* (and at this stage of the argument that's all it is) there is a notion of responsible agency and a justifiable system of praising and blaming that can adequately function without an incompatibilist metaphysics, then it looks like this is all we need to be revisionists, instead of eliminativists or nihilists about this notion of free will. And (once again) *if* we locate such a notion, then we might find ourselves in the fortunate position of not missing those libertarian elements of our self-conception, anyway.

On the approach I am suggesting, in the face of libertarianism's implausibility, the main question before us is whether we still have reason to go on roughly as before. If we do not have reason to go on roughly as before, then eliminativism becomes more plausible. If however we can locate some reasons to continue as before, then revisionism is surely the more credible option.

5. How to solve the problem

Suppose you decided to take seriously the thought that free will might be different than we sometimes tend to think, but that it should be the sort of thing intimately tied to the business of moralized praising and blaming. How might you go about trying to construct an account of what that might be?

There are at least two options available to you. I call them *repurposing revisionism* and *systematic revisionism*. The first option is the easiest: take any traditional compatibilist theory you like, declare that it is only prescriptive and not diagnostic, and *voilà!,* you have a revisionist theory. The virtue of this approach is that you have some already existing theories at your disposal. The downside is that those theories tend to have been constructed in a context where the concern for intuitiveness played an impor-

tant role in the construction of the theory. Most compatibilist theories have not started with the question of what could justify our practices of praising and blaming (and the typical web of connected judgments and attitudes), but have instead begun from trying to capture our intuitive judgments and ordinary patterns of the ways we in fact praise and blame. So, you might worry that these accounts will contain elements in them that were developed out of a concern for commitments or intuitions that the revisionist can reject.

Systematic revisionism constitutes a more demanding but also more appealing alternative. On this approach, the strategy is to begin with a picture of what we want a theory of moral responsibility to *do*. On the model I favor, the goal of the theory is to identify features of agency that can play an appropriate role in explaining the justification of our practices of praising and blaming and that helps explain how familiar judgments and attitudes about freedom and responsibility make sense in a framework that minimizes *ad hoc* metaphysical commitments.

The last two ideas can be expressed in terms of commitments to *normative adequacy* and to *naturalistic plausibility* (or perhaps less dogmatically, *scientific plausibility*). Roughly, the idea is that we want our theory of free will to be beholden to three things: (1) some conceptual role (in this case: a distinctive agential feature whose presence and operation licenses moralized praising and blaming); (2) explanatory and justifying tasks connected to the particular conceptual role identified by the account (in this case: an account of why this power would be the sort of thing that licenses praise and blame); and (3) a picture of the capacity or power that constitutes free will that does not put us at odds with our epistemically best pictures of the world (in this case: no libertarianism, and the aspiration to not run afoul of our scientific understanding of how agency operates). So, proposed departures from our ordinary ways of thinking won't be unprincipled if they are responsive to these concerns.

There are several advantages for any account developed along these lines. First, it would have the following advantage over conventional compatibilist accounts: it need not deny the existence and deep-rootedness of incompatibilist commitments among ordinary folks. Second, it needn't be committed to armchair neuroscience, old-fashioned speculative meta-

physics, or other troublesome endeavors that arise when philosophers attempt to sketch pictures that commit us to very particular and otherwise unmotivated views about how future science will unfold. Third, it would provide us with a principled explanation for why the identified powers matter, how it is that these things—despite their running afoul of some of received convictions—are the sorts of things that properly underpin attributions of freedom and responsibility.

Of course, these advantages are conditional on actually generating a revisionist account that satisfies the afore-mentioned constraints on revisionist theory-building. So, more needs to be said.

6. Outlines of a solution

Now is the part of the story in which I downplay the demand for an account that would provide the antecedent of the extravagant conditional claims in the previous section. I do this for good reason: it is far beyond the scope of the present discussion to present a worked out account of how a systematically revisionist, prescriptively compatibilist account of free will might go, for such an account necessarily involves thorny details far beyond the scope of a single chapter (Vargas 2013). Moreover, as I emphasized at the outset, the aim of this particular essay is less to convince you to embrace my particular account than it is to make plausible a general approach that might intrigue those who remain unhappy with the more familiar options.

Although in what remains of this essay I say a bit about my preferred approach, I do not mean to suggest that there are not other revisionist accounts that could be given. Indeed, revisionism constitutes a class of theories, of which there are many possible instances. So, what follows should not be taken to be the definitive statement of what any revisionist account should look like. On the contrary, it would be best if we were in a position where we could select among multiple competing revisionist accounts, weighing contrasting advantages and disadvantages of theories, liberated from the shackles of our otiose or problematic intuitions.

Of course, I'd be delighted if the brief sketch that follows were sufficient to convince you to follow up on the details, but I'll be happy

enough if I make plausible the thought that we are tantalizingly close to a solution to the problem of free will.

Let's start with the question of what, if anything, could justify praise and blame—apart from our being libertarian agents. This is a reasonable place to start precisely because we have yoked our picture of free will to the role in plays in moral responsibility. Given this framework, here's an appealing answer: praise and blame would be justified if they played some role in attaining some other end whose value is substantial and clear. Here's one such end: enhancing our agency, in particular, our capacity to recognize and respond to moral considerations. To the extent to which our practices of moral responsibility, including praise and blame, play some appropriate role in supporting and enhancing the ongoing success of such agency (both in terms of responsiveness to moral considerations and in terms of the scope of context of actions in which such agency operates sufficiently well), we have a plausible justification for praising and blaming.

I'm glossing over a number of important details here. For example, on the picture I favor, the relationship between praising and blaming practices and the modest teleology that structures those practices is two-tiered: the ground level practices are not themselves goal structured in their content or how they are ordinarily regarded—the teleology is only in the justification of those non-consequentialist elements. Moreover, the standards of praise and blame have their own internal logic, one that roughly tracks a concern for quality of will.

Here's what all of this has to do with free will: free will is the capacity we have to recognize and respond to moral considerations. This is a picture on which the possession of free will partly explains *why* we can be morally responsible. Without our having free will, responsibility practices lose their point, for there would be no agency of the relevant sort to enhance. It is also a picture that explains why free will is distinctively valuable. It is in virtue of having free will that we are the sorts of morally significant creatures that we are.

Of course, free will in the absence of a suite of standard agential capacities (for example, the ability to foresee consequences of action, to reason instrumentally, to settle on plans that get filled in over time, and so

on) will be uninteresting. So there are some background presumptions about other agential capacities, but assuming their presence, then free will possesses a kind of explanatory power in understanding the normatively structure social practices that fill our daily lives.

Important to my account is the idea that the capacities to recognize and respond to moral considerations are multiply realizable. Indeed, their nature and structure are not stable across contexts. Bracketing some complexities, the picture is one on which the metaphysical realizers of free will vary: what counts as as a sufficient degree of responsiveness to moral considerations is not the same in all cases. This variation is partly a function of the justifying teleology of the responsibility system and the variable psychological constraints we operate under. Given that circumstances plausibly structure the capacities we have, a practicable system of normatively ideal standards for considerations-responsiveness will vary across contexts.

This picture also entails that our having free will is not an all-or-nothing affair, and that it is the kind of thing that we might grow to have in some contexts while continuing to lack it in other contexts. One byproduct of this picture is an original way of rewriting talk of alternative capacities. It is one on which the relevant notion of alternative possibilities is not to be extracted from the brute features of the metaphysics of agents, but instead from a complex relation consisting of the metaphysics of the agent and the agent's position in a web of normatively structured interests and practices.

What this means is that our having free will is not an intrinsic feature about us, but a partly relational notion. In turn, this means that we can ask questions about the circumstances of action, and whether our local moral ecology is conducive to our being responsible agents. The questions of whether, and how, and to what extent we build such circumstances are difficult questions that I have not attempted to answer. If I am right, the deepest problem that threatens our having free will is not a matter of high metaphysics, but rather the contexts in which we exercise our agency and the political challenges of structuring our environments to better support responsible agency.

My hope is that the preceding sketch is sufficiently provocative to make plausible the idea that we have considerable resources available to us, if we wish to walk down a revisionist path. If I am right, we have free will, and it is compatible with what we know about our agency and the world. It is not the notion of free will we started off looking for, but it is a notion that leads us away from responsibility nihilism.

7. Is revisionism just another cheat?

Among the many memorable condemnations of compatibilism, one of the best is Kant's remark that it is a "wretched subterfuge." Given that the account I have suggested is prescriptively compatibilist (although *not* diagnostically compatibilist), we might wonder whether it amounts to yet another wretched subterfuge.

In reply: even if my version of revisionism is wretched, it is most certainly not a subterfuge. My account begins with the idea that an adequate theory of free will costs us something. It costs us a piece of our self-conception, that part of our self-conception that sees us distinguished from the ordinary causal order of the universe, possessed of a unique ability to screen off the past and initiate new chains of causation disconnected from prior facts. What free will skeptics would like you to think is that this concession means that we lack free will and moral responsibility, just as some 19th century atheists would have had you believe that the death of God entailed that everything is permitted.

Of course, it might turn out that we lack free will for some other reason, apart from our being part of the natural causal order. And, it might turn out that morality is bunk not because God doesn't exist but because there is no way to, say, draw a distinction between moral norms and merely local cultural conventions. Nevertheless, the mere fact that things are not as they first seem is not proof that they do not exist. It might only be proof that some more thinking is required before we understand what we are talking about.

References

Fischer, John Martin. 2006. *My Way: Essays on Moral Responsibility*. Oxford: New York.

Franklin, Christopher. Forthcoming. The Scientific Plausibility of Libertarianism.

Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. 2006. Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research* 73 (1): 28-53.

Nichols, Shaun, and Joshua Knobe. 2007. Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Nous* 41 (4): 663-85.

Nietzsche, Friedrich Wilhelm. 1966. *Beyond Good and Evil*. Translated by Kaufmann, Walter. New York: Vintage Books.

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

Pereboom, Derk. 2006. Kant on Transcendental Freedom. *Philosophy and Phenomenological Research* LXMIII (3): 537-67.

Roskies, Adina, and Shaun Nichols. 2008. Bringing Responsibility Down to Earth. *Journal of Philosophy* 105 (7): 371-88.

Sarkissian, Hagop, Amita Chatterjee, Felipe De Brigard, Joshua Knobe, Shaun Nichols, and Smita Sirker. 2010. Is Belief in Free Will a Cultural Universal? *Mind and Language* 25 (3): 346-58.

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

Vargas, Manuel. 2013. *Building Better Beings: A Theory of Moral Responsibility*. Oxford, U.K.: Oxford University Press.

Woolfolk, Robert L., John Doris, and John Darley. 2006. Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition* 100 283-401.