The Revisionist's Guide to Responsibility

(Forthcoming in *Philosophical Studies*)

Manuel Vargas

Revisionism in the theory of moral responsibility is, roughly, the idea that some aspect of our responsibility practices, attitudes, or concept is in need of revision. In this paper, I argue that (1) in spite of being an increasingly prevalent thread in discussions of moral responsibility, revisionism is poorly understood, (2) the limited critical discussion there has been of it does not reflect the complexities and nuances of revisionist theories, and (3) at least one species of revisionism— moderate revisionism— has some advantages over conventional compatibilist and incompatibilist theories. If I am right, one result is that the outcome of prominent debates about the compatibility (or not) of determinism and our commonsense thinking about moral responsibility may be less crucial than they seem.

1. Revisionism and its critics

Though it is more frequently hinted at, rather than systematically pursed, the increasing frequency of revisionist talk about free will and moral responsibility is striking. The best-recognized place for it is a wing of compatibilism (championed by Dennett) that invokes considerations about "the varieties of free will worth wanting."[1] Within incompatibilism, though, there is also a well-established tradition of revisionism: (in)famously, hard determinists usually argue that significant, even radical revisions in our ordinary concepts and practices are required because (so they believe) determinism is true.[2] Various non-standard views about free will, including Fischer and Ravizza's "semi-compatibilism" might also be taken as revisionist in one or another fashion.[3] Revisionism (and what should count as an instance of it) is thus an issue for a large part of philosophical discourse about responsibility and free will.

What few systematic discussions there have been of revisionism tend to be strongly critical, though in some cases this is to be expected.[4] For example, hard determinism, like any skeptical position, is likely to always have vigorous critics. What *is* notable is that much of the resistance seems directed at milder forms of revisionism, forms that we might broadly identify as "deflationist." Incompatibilists and compatibilists alike have attacked deflationist forms of revisionism. For instance, it is intended as a criticism when Timothy O'Connor describes Robert Kane's form of event causal libertarianism (a version of libertarianism that does not rely on irreducible agent causation) as "somewhat deflationary."[5] Colin McGinn, a compatibilist of sorts, has criticized all deflationist forms of revisionism about free will and moral responsibility as well, and arch-incompatibilist Peter van Inwagen has endorsed McGinn's criticisms.[6] What makes this resistance remarkable, even puzzling, is why it is happening now. As O'Connor himself has noted, naturalism carries with it a tendency towards deflationism about the traditional topics of philosophy.[7] In our current climate of casual naturalism, it should therefore seem surprising if fairly tame revisions are criticized simply for being revisionist. It would be even more surprising if these forms of revisionism had received no serious, systematic defense. But, this is precisely what has happened, and it is precisely the problem I aim to address by providing a revisionist's guide to the theory of responsibility.

Interestingly, criticism of revisionism has been far more systematic than the pursuit of revisionism. Even theories with explicitly revisionist elements usually downplay it, adopting it as a matter of last resort.[8] I believe that much of the hesitancy and dissatisfaction with milder forms of revisionism stems from a lack of understanding on all sides about what this sort of revisionism entails and how it is different than other forms of revisionism. For instance, reluctance to allow for revisionism about responsibility can be tied to the worry that *any* revisionism about

responsibility would entail consequences similar to those advocated by hard determinists. Also, one might doubt that there is any standpoint from which to anchor or restrain proposed revisions should we turn out to be incapable of fully vindicating our commonsense concepts of responsible agency. In the absence of any attempt to develop a systematically revisionist theory, resistance to revisionism may simply reflect reasonable resistance to *ad hoc* or apparently unprincipled approaches to the theory of moral responsibility. Without a clear view of what revisionism entails, suspicion about its adoption or pursuit may be justified.

In this paper I aim to undermine the broad and ready skepticism facing revisionist approaches by articulating a clear picture of what forms revisionism can take and what those forms entail. Unsurprisingly, not all revisionisms are created equal. One particularly promising version of revisionism—*moderate revisionism*— appears to be immune to many of the worries that seem to fuel hesitancy and resistance to revisionist approaches. By combining the folk conceptual analysis of incompatibilism with the metaphysical minimalism of compatibilism, it avoids many of the strongest objections leveled at both incompatibilism and compatibilism while picking up many of their chief advantages.

In light of the above conclusions, I will argue that there may be good reasons to prefer a principled revisionist theory to standard non-revisionist approaches. Moreover, revisionism may free us from at least some of the difficulties that plague contemporary debates about free will and determinism.

In what follows, my focus is on responsibility and free will, the latter to the extent that it matters for the former. This is, after all, a revisionist's guide to *responsibility*. By that I mean that I will treat *free will* as a technical term referring to the freedom condition of moral responsibility. Theories that distinguish between these concepts (e.g., Fischer and Ravizza's

semi-compatibilism) will be understood in terms of the analysis of responsibility. This gives my discussion a very particular slant, but it is not meant as a substantive claim that restricts concerns about free will to concerns about moral responsibility. Though many of the details would surely change, one could rewrite much of this discussion to apply to some notion of free will other than the freedom condition on moral responsibility.

2. What is revisionism?

I take it that philosophical concern with responsibility is bound up with at least three interrelated but distinct elements. First, there are the characteristic psychological dispositions or attitudes associated with responsibility (roughly, what Strawson called the "reactive emotions," e.g., resentment and gratitude). Second, there are the practices associated with responsibility (e.g., sanctioning, rewarding, praising, and blaming). Finally, there is the cluster of beliefs, judgments, or concepts about responsibility (including beliefs about the conditions for appropriate ascription of responsibility) that we might broadly call the folk concept of responsibility.

Though our concerns are spread over the elements just mentioned, I take that there are a number of particular questions one might attempt to answer with a theory of responsibility. One question is what we might call *metaphysical* in the broad sense: "What is the nature of responsibility?" Other questions include: "What do we think about responsibility?" and "What *should* we think about responsibility?" In principle, we might offer different answers to each of these questions, though presumably they will overlap in various ways.

Revisionism emerges out of a difference between the projects invited by the last two questions. One project is *diagnostic*, for it attempts to give a diagnosis of our commonsense

reflections regarding responsibility. A diagnostic account of responsibility would be concerned to reflect the facts about our concept of moral responsibility and its conditions of application. The second project is something we might call *prescriptive*, for it aims at generating a theory that can guide our thinking about (and practice of) responsibility. In other words, it tells us what we should think and do.

A typical theory of responsibility will contain elements of some or all of the above elements. And, diagnostic and prescriptive projects doubtlessly have important connections to one another and to the more general question of the nature of responsibility. For many, an ideal result would be the discovery that we should and in fact do think in some way mirrors what is the case about responsibility itself. However, we need not suppose that such a felicitous alignment of ontology, normativity, and actual practices has occurred.

This possibility, the one that allows for a difference in diagnostic and prescriptive projects, captures the idea of revisionism that is central to this paper. A theory is for our purposes *paradigmatically revisionist* if it prescribes something other than what it diagnoses. In contrast central case of a conventional, *non*-revisionist theory is one where the diagnostic and prescriptive aspects of the theory do not bifurcate, treating diagnostic and prescriptive aspects of the theory in a unified way. Thus, to determine whether a theory is revisionist or not in the paradigmatic sense, we need only determine whether its diagnosis and prescription are the same or different.

There are other kinds of theories that could be reasonably characterized as revisionist, though not in our paradigmatic sense. It may help to clarify matters to distinguish them from one another and from paradigmatic revisionism. One alternative variety of revisionism concerns departures from accurately characterizing the nature of responsibility— something we might call *de re* revisionism. A theory is *de re* revisionist if it does not accurately reflect the nature of

responsibility, whatever that may be. For example, if it turns out that our responsibility-characteristic practices are, as compatibilists maintain, compatible with determinism, then a theory that holds that they are not (e.g. a libertarian theory) is *de re* revisionist about that our responsibility-characteristic practices. Or, if agent causalists are right, event causal libertarians and compatibilists will count as *de re* revisionists. Until we find a way to decisively determine the nature of responsibility, attributions of *de re* revisionism will be a largely partisan endeavor. Still, we can say that all theories that aim to correctly characterize the nature of responsibility and get it wrong will be *de re* revisionist about what they get wrong. Naturally, few conventional theories of responsibility are likely to set out to be *de re* revisionist, for to be described as such will usually amount to a polite way of saying the theory is false, bad, or otherwise unsatisfactory.[9]

Another way in which a theory might count as revisionist is if it prescribes responsibility-characteristic practices, attitudes, beliefs, or our construals of them, that are different than the ones we in fact have. This form of revisionism can be called *de facto* revisionism. Our ability to determine whether a theory is *de facto* revisionist turns on the extent to which we have correctly characterized responsibility practices, attitudes, and beliefs. To the degree to which these are unsettled or disputed, categorization of *de facto* revisionist theories will be contentious.

De facto revisionism cuts across paradigmatic revisionism, and thus some instances of *de facto* revisionism will also count as instances of paradigmatic revisionism, and vice-versa. Indeed, the success of most paradigmatic revisionisms will partly hinge on also being *de facto* revisionist. A paradigmatic *de facto* revisionist theory is one in which the prescriptive aspect of the proposed theory differs from both its diagnostic aspect and the facts about our construals of our responsibility-characteristic practices, attitudes, and beliefs, or those elements themselves.[10]

Other forms of paradigmatic revisionism will not count as *de facto* revisionism, however. For example, a revisionist theory which misdiagnoses our folk concept of responsibility, and as consequences prescribes changes that are falsely believed to be departures from how we think about responsibility would not be *de facto* revisionist. Its *de facto* non-revisionism would simply reflect the fact that it prescriptive element does not depart from how we actually think about responsibility. [11]

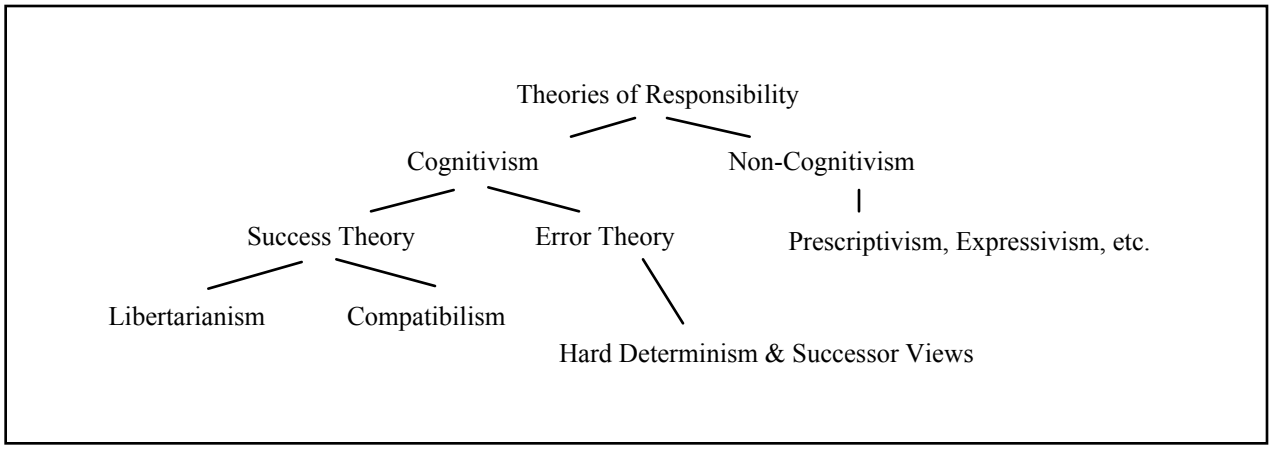
In what follows, I focus on paradigmatic revisionism. Consequently, further uses of the unmodified term 'revisionist' should be taken to refer to paradigmatic revisionism. This sort of revisionism has the benefit of being comparatively easy to identify: any theory that holds that our thinking is one way but should be another will be an instance of paradigmatic revisionism. Of course, revisionism may be inadvertent, unnoticed, or unappreciated by its author(s). A theory could have prescriptive implications that an author has missed or failed to fully draw out— *de dicto* non-revisionism, if you like. Moreover, there might be systematic ambiguity in a proposal that makes some cases impossible to determine whether it is revisionist or not. In spite of the complexities involved in making these determinations (see §6 for some of them), these challenges are minor compared to those involved in making determinations about *de re* and *de facto* revisionism. We would need a true and complete account of the nature of responsibility and a true and complete account of our responsibility-characteristic practices, attitudes, and beliefs to decisively settle whether any particular theory counts as *de re* or *de facto* revisionist. Of course, for many philosophers *de re* issues are the central concern for a theory of responsibility. But, as we will see in sections 6 and 7, *de re* revisionism is less of a concern than we might suppose.

In the next section, I take up the relationship between paradigmatically revisionist theories and more conventional approaches to responsibility.

3. The metaethics of responsibility

It is customary to mark out the available terrain for theories of responsibility as answers to the compatibility question, the question of whether moral responsibility is compatible with determinism. For present purposes, it is more useful to begin by thinking in terms of how conventional categories in the theory of responsibility map on to a framework borrowed from metaethics.[12]

Consider the following taxonomy of theories, which I will call R:



As diagram R (fig. 1) indicates, theories of responsibility can be cognitivist or non-cognitivist. This division captures a distinction between theories that take talk of responsibility to have truth-conditions (cognitivist theories) and theories that deny this (non-cognitivist theories).

Since it is not customary to discuss theories of free will and responsibility in terms of cognitivism and non-cognitivism, it is not always obvious when a theory is meant to be non-cognitivist. For example, though it is tempting to read Double' as a non-cognitivist, Pereboom reports that he is not.[13] There are some clearer cases of non-cognitivism: P.F. Strawson's classic "Freedom and Resentment" is one, and Honderich's *A Theory of Determinism* is another.[14]

However, given that most theories of responsibility seem to be cognitivist and the main features of revisionism— typically, prescriptions for change in our ordinary moral thinking and practices, and what follows— can be reproduced across the cognitivism/non-cognitivism divide, for ease of exposition I will focus on cognitivist theories. Though some of the details would be different, much of what I say about cognitivist approaches can be easily reproduced on the non-cognitivist side.

However you choose to spell out the truth conditions of responsibility, cognitivist theories are either success theories or error theories.[15] The former hold that at least sometimes ascriptions of responsibility are true whereas the latter holds that ascriptions of responsibility are never true. There are two main species of success theory, which together capture the bulk of mainstream theorizing about responsibility. The first strand is libertarianism. These are the familiar theories that hold that responsibility and determinism are incompatible, but that we at least sometimes satisfy conditions for responsibility. The second strand of success theories is compatibilist. These theories hold that responsibility is compatible with determinism, and most contemporary versions hold that the question of determinism (or naturalism more generally) is altogether irrelevant to the question of responsibility.[16]

The best-known examples of error theories are hard determinist theories. These are accounts that maintain that responsibility is incompatible with determinism, and that since determinism is true, no one is ever responsible. Hard determinists these days are a rare breed, but a number of philosopher have taken to defending nearby relatives that we might, following Derk Pereboom, call "hard incompatibilist" theories of responsibility. These are views that hold that responsibility is incompatible with both determinism and indeterminism. As most of these theories are presented, they amount to *eliminativist* views of at least the concept of responsibility.

That is, these views hold that talk of responsibility ought to be eliminated from a suitably rigorous discourse.

4. Revisionism: some varieties

Though perhaps the bulk of compatibilist and incompatibilist theories are not revisionist in any significant way, there are at least three kinds of paradigmatic revisionism available to philosophers of moral responsibility. Call these revisionisms 'weak', 'moderate', and 'strong'.

A *weak revisionist* theory revises beliefs about the various elements of concern for a theory of responsibility. According to weak revisionism, the concept of responsibility and the associated practices and attitudes do not themselves require revision, but our understanding of one or more of these does. As weak revisionists see it, we have come to misunderstand our own concept, practices, or attitudes, as they actually exist. Usually, our misunderstanding is a result of some confusion introduced by philosophical speculation, or some other correctable defect of cognition that keeps us from seeing what we *really* believe, mean, feel, or do. But if we just understand things properly, all that needs revising is our understanding of things, not the concept, practices, or attitudes, *per se*.

A wide range of historical and contemporary compatibilist theories of responsibility will count as instances of weak revisionism, from early 20th century compatibilists like Moore and Schlick to contemporary (semi-) compatibilists such as Fischer and Ravizza. Perhaps the best examples are conditional analysis compatibilists (or "conditionalists") in the tradition of G.E. Moore. They maintain that people may ordinarily say, and maybe even believe (usually under the spell of certain philosophers), that responsibility requires that one can do otherwise, understood in some "categorical" or incompatibilist sense. But, the conditionalists maintain, these people

simply misunderstand the responsibility-relevant notion of 'can' or 'control' or 'power' or what have you. If we just understand the relevant concept properly, we can see that our concept (or practices, or attitudes) is not actually incompatible with determinism. Hence, it is not our concept, practices, or attitudes that is in need of revision. Rather, it is our understanding of them that needs revision.[17]

In a similar vein, Fischer and Ravizza argue that our concept of responsibility-relevant control is devoid of the metaphysics typically insisted on by incompatibilists about responsibility, though they admit that there are intuitions that initially suggest otherwise. On their view, what needs revision is not the concept of responsibility itself, but only our mistaken interpretation that its conditions of application include the presence of a "metaphysically robust" incompatibilist free will. "Frankfurt-style examples" purportedly illustrate the error of our perhaps natural inclination to think in incompatibilist terms.[18] According to Fischer and Ravizza, such examples show that even though we ordinarily suppose that responsibility requires alternative possibilities (which they maintain would be incompatible with determinism), our thinking about responsibility– contra our typical self-assessment— really is compatible with there being no alternative possibilities.[19] Since their account has separate diagnostic and prescriptive elements, it is revisionist. And, since the revision it recommends is one in self-conception (not in the concept itself, its conditions of application, or the associated practices), their theory is a form of weak revisionism. (Which, of course, is not to suggest that it is weak in some other sense!)

In contrast to weak revisionism, *strong revisionism* maintains that our concept, practices, or attitudes themselves are in need of elimination. Where weak revisionism merely maintains that we need to modify our understanding of responsibility, strong revisionism argues that we

must dispose of some or all of the main elements addressed by a theory of responsibility. Strong revisionism is a kind of eliminativism of the sort traditionally associated with hard determinists at least as far back as Spinoza, and more recently, hard incompatibilists such as Pereboom and Galen Strawson.[20] For these "pessimistic" incompatibilists, strong revision is usually taken to be a necessary implication of skepticism about responsibility. If we are not responsible, so the argument goes, then some to all of the practices, attitudes, or clusters of beliefs we have about moral responsibility must go as well.[21]

One could protest that strong revisionism is not a form of revisionism at all, for the change it recommends is elimination, not revision. This concern is merely terminological. If you prefer, you may think of the categories in the following way— various theories maintain that we need to make changes in our beliefs, practices, or attitudes characteristic of moral responsibility. Those changes come on a spectrum from minor to major. On one end are minor changes such as clarifying relatively simple linguistic and conceptual confusions. On the other end are major changes that may include the elimination of various practices, concepts, and psychological states. Changes on one end I am calling weak, changes on the other I am calling strong.

Between the poles of weak and strong revisions, there is a wide range of views that we can call *moderate revisionism*. When applied to the folk concept of responsibility, moderate revisionism is the idea that the folk concept of responsibility is inadequate until it has been modified in some way. Unlike strong revisionism, moderate revisionism's revision does not involve straightforward elimination of the concept, practices, or attitudes characteristic of responsibility. Rather, it amounts to a "pruning" of that element. This pruning may itself involve eliminating some aspect of the considered element but it does not require elimination of the entire element. For example, moderate revisionism regarding attitudes might counsel eliminating

retributivist attitudes from our network of responsibility-characteristic attitudes without counseling wholescale elimination of all of our responsibility-characteristic attitudes. Or, it might recommend modifying some of our responsibility practices without abandoning them altogether. Thus, in some suitably broad sense, we might think of moderate revisionism as a form of deflationism.[22]

5. Moderate revisionisms

Moderate revisionism introduces a puzzle: does it correspond with any positions on R (fig. 1)? Not without some additions. This is because strands of moderate revisionism seem to map on to both compatibilism and incompatibilism in very distinctive ways. To see how, it helps to consider carefully what moderate revisionism is committed to, at least in one form.

Consider moderate revisionism about the *concept* of responsibility. It may turn out that any adequate revision of the concept entails revisions in practices or associated attitudes, but for present purposes we can ignore this possibility. What moderate (conceptual) revisionists agree on is that the concept of responsibility stands in need of some revision. These revisionists might be thought of as "folk conceptual error theorists," for they agree that there is some error in ordinary thinking about responsibility.

At this point, issues of semantics come into play. Consider the following, which we can call the Familiar Argument: “If it turns out that our commonsense thinking about responsibility (or free will) is deeply flawed, then we are never suitably entitled to hold people responsible. This result is unacceptable, obviously false, or patently absurd. So, we have to show that commonsense can be vindicated.” If we accept the Familiar Argument, then what I have been calling moderate revisionism seems to be a species of incompatibilism. However, we need not

accept the slide that starts with the failure to vindicate our folk concept and ends with the claim that we are never responsible. We could hold that our error-ridden *concept* of responsibility shows us little, if anything, about responsibility itself. For instance, we might hold a kind of causal theory of reference where participating in the right causal chain allows us to refer to responsibility even if most of our beliefs about it are mistaken. Alternately, we could have a theory (perhaps something like Searle's weighted cluster of beliefs), where our beliefs about responsibility *do* fix reference. In this case, we might contend that a folk conceptual error does entail an error about the property. The point is that identification of a conceptual error (i.e., the folk conceptual error theory) does not, by itself, commit us to thinking there is no (instantiated property of) moral responsibility.[23]

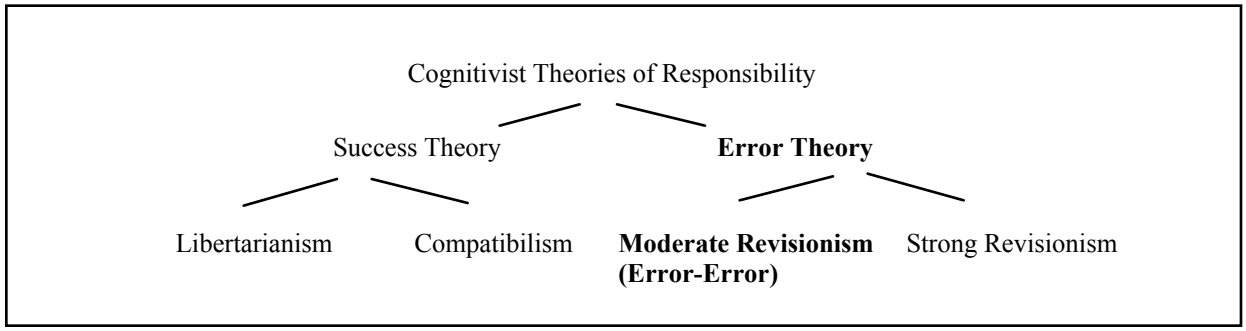
It is worth noting that though few, if any, philosophers have stated the Familiar Argument so baldly, it does seem to lurk in the background of much of the contemporary literature. Few philosophers bother to argue for *why* our ordinary intuitions must be vindicated. Rather, most start with the issue of whether incompatibilism or compatibilism about moral responsibility captures our intuitions. Considerations that generally favor conceptual conservatism may provide some support for this approach, but it also may be part of a broader pattern of conventional theories failing to distinguish between the various philosophical projects that I mentioned earlier: diagnostic, prescriptive, and metaphysical. In conventional non-revisionist theories, diagnostic and prescriptive projects are collapsed: the account of our current responsibility practices is taken to be both descriptive of our current concept and associated practices and attitudes and prescriptive for how such characteristic practices should be. The Familiar Argument takes this one step further by supposing that our thinking about responsibility must accurately reflect the facts about the world, thus collapsing all three projects— diagnostic, prescriptive, and

metaphysical— into a single account. The attraction of this singular approach is obvious: it turns out that what we already believe both describes the way things are and the way they ought to be. The point here is that the revisionist need not share this remarkable belief in such a pre-established harmony.

In sum, the difference between a folk conceptual error theory and more traditional property error theories show that moderate conceptual revisionism comes in at least two main types: "folk conceptual error/property error theories" and "folk conceptual error/property success theories." For short, we can call them 'error-error theories' and 'error-success theories'.

Error-Error Theories

Consider the error-error theory. As the name makes clear, accounts of this type should be placed on the error theory branch of diagram R (fig. 2).



But, by definition, a moderately revisionist theory is not committed to wholescale eliminativism of the sort associated with traditional hard determinism. If error-error moderate revisionism is a genuine alternative to hard determinism, it suggests that philosophers frequently have been mistaken in supposing that all error theories of responsibility entail eliminativism of the sort characteristic of hard determinism.

There is a not uncommon picture of eliminativism (strong revisionism) about responsibility that holds that if responsibility judgments or ascriptions are systematically false, then all the characteristic features of responsibility must be abandoned. The assumption is that eliminativism about the concept entails eliminativism about (many or most of) the associated practices and attitudes. While it is not clear whether or how many incompatibilists have actually held this sort of view, it nevertheless plays a significant role in the economy of free will debates. Compatibilists get considerable leverage out of inviting readers to imagine whether incompatibilists would really stop holding people responsible if physicists one day issued a declaration affirming the truth of universal determinism.[24] The implausibility of incompatibilists doing so can be taken as evidence against the plausibility of incompatibilism in general and error theories in particular. Of course, eliminativists have attempted to explain how this outcome is not deeply problematic, and how it might even be desirable.[25] Whatever the case may be it is clear that thoroughgoing eliminativists face an uphill battle trying to defuse the all-too-common view that error-theoretic views about moral responsibility amount to a *reductio ad absurdum*.

Moderate revisionist error theories provide a less dramatic, perhaps more reasonable alternative. Sophisticated revisionists need not claim that moderate revisionism or strong revisionism about the concept of responsibility entails the same about relevant practices and attitudes. Since we typically have diverse concerns in a theory of responsibility, there is (in principle) room to be moderately revisionist or deflationary about some aspects, eliminativist about others, and conservative about the rest. One might thus endorse deflationism about the concept and some practices while adopting conservatism about attitudes. This would allow the revisionist to say that the concept of responsibility needs to be changed without forcing the

resulting theory to endorse a heavy-handed eliminativism about the involved psychology. Revisionism can thus vary between categories.

Revisionism can also vary within categories. Consider the various possible targets for purely conceptual revisionism: epistemic conditions, or the kinds of things a responsible agent must know; the freedom condition, or the kind of freedom a responsible agent must have; the ultimacy condition, or some notion that the considered act or state of affairs is ultimately up to the agent; various considerations about rationality; and presumably, a capacity for consciousness. In principle, any of these things are open to deflation or elimination, in any combination. So, for example, the error theory might opt for strong revisionism about ultimacy and moderate revisionism about the freedom condition.

Given the variety of available revisionisms, any theory that capitalizes on these complexities will need to be cautious about specifying what is to be revised. This is important not just for the sake of clarity, but also because revisionists face unique burdens. In particular, moderate revisionists must not only show that the revision is compatible with the favored worldview, but that the revision is plausible and normatively warranted.

The plausibility constraint on revisionist theories is easy enough to understand: a revision that requires something that is not psychologically possible or socially implausible is likely to be and perhaps ought to be rejected.[26] Similarly, revisionisms that rely on highly speculative or largely implausible accounts of agency will fare worse than revisionism that do not rely on dubious pictures of agency.

Perhaps the most complicated task facing the error-error revisionist will be to specify the kind of warrant that guides the proposed revision. That is, the moderate revisionist will need to say what it is that justifies changes in our beliefs, attitudes, or practices, and how, given those

changes, the revised thing still merits the name 'responsibility'. A likely initial answer will focus on the way the revision preserves the bulk of responsibility-characteristic attitudes and practices. But a deeper problem lurks. Since responsibility is an almost fundamental moral notion, a revised account of it will need to specify a non-responsibility-dependent position from which to revise the concept. The details of revisionist metaphysics and semantics will vary by revisionist. For instance, one revisionist might maintain that the successful re-anchoring of our responsibility practices and talk is made possible by there being a property such that if we were to talk about it, and to use it as a foundation for our responsibility-characteristic practices, our characteristic attitudes, judgments, and practices would be well-founded. Another might focus on arguments to the effect that the bulk of our practices and attitudes can be intersubjectively justified on multiple grounds, each independent of some external metaphysical fact of responsibility. In advance of the development of individual error-error moderately revisionist theories, it is difficult to predict what the most appealing versions will be. However it goes, though, we can expect that as a *moral* concept, whatever revisions are proposed have to be justifiable in the same way as our other moral commitments, and coherent with them. In time, the revisionist is likely to make the reconstruction of responsibility look more like a piece of ethics than a piece of metaphysics.

Compare the case of the conditionalist (a weak revisionist). There, the warrant for revision is straightforward: the actual meaning of responsibility. Weak revision of attitudes and practices would presumably have similar warrants generated by our actual attitudes and practices. One way generating the relevant warrant is to get it from a higher epistemic authority. Such warrant is often given by empirical studies. Thus, certain sociological and psychological theories of responsibility might count as weakly revisionist in a less contentious way than philosophical disputes about, e.g., the meaning of 'can'. To the extent that such empirical

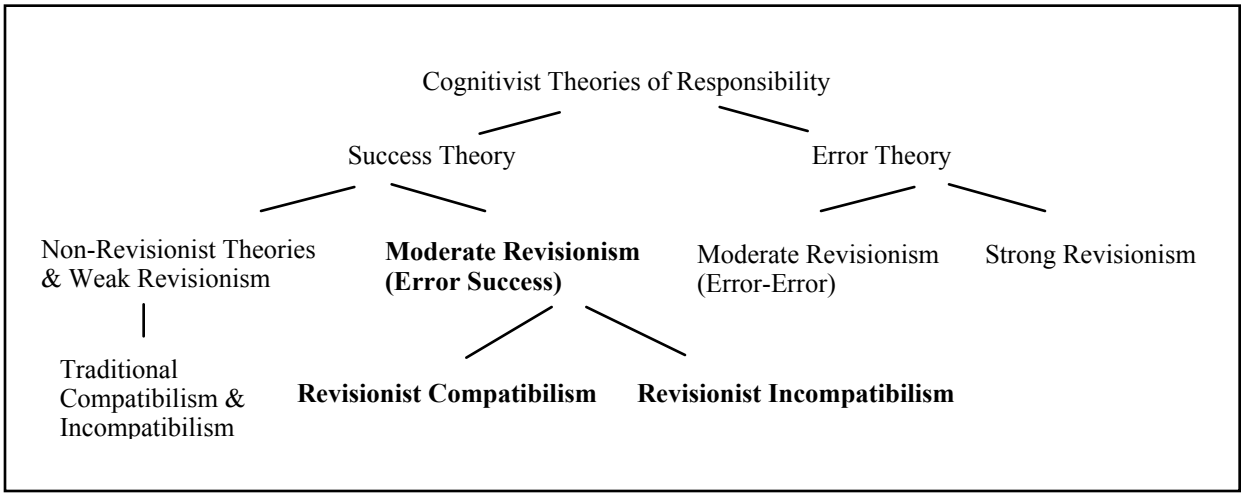
accounts alter our everyday understanding of, say the practices or attitudes characteristic of responsibility, such theories will count as weakly revisionist theories of responsibility. Of course, these accounts are not immune to many of the normative issues raised by more vigorously revisionist approaches. Even if we accept one or another account of actual usage, we can still raise questions about whether we want to preserve our current usage.

Still, for incompatibilists who believe that a great deal of restraint ought to be exercised when making changes to our moral practices in light of metaphysics, a viable form of moderate revisionism will be seen as something of an advance. Where traditional hard determinists conclude that responsibility and the things characteristic of it must be abandoned, moderate revisionists contend that with some revision, our practices and attitudes can continue, indeed, in better epistemic and normative shape than before. This point holds even for incompatibilists who are libertarians. If libertarianism is empirically falsified— something most libertarians concede is a possibility— moderate revisionism provides a plausible, normatively adequate account that accommodates the intuitions that motivate incompatibilism.

There are obviously a variety of error-error revisionisms available, and their burdens will differ depending on just what is being revised and why. The lesson for us, though, is that we need not perpetuate the mistake of assuming that any error theory about responsibility entails wholesale eliminativism about the main elements of a theory of responsibility.

Error-Success Theories

Let us now consider the other moderately revisionist possibility, the error-success theory (see fig. 3).



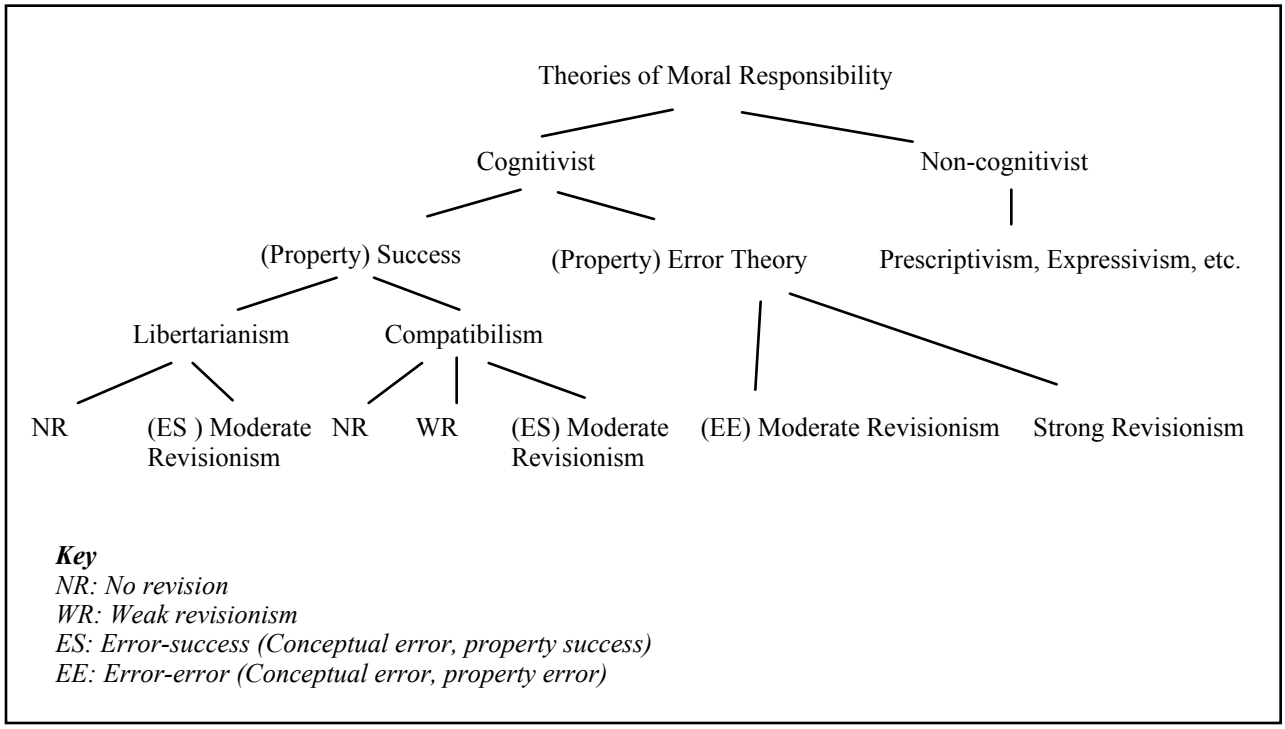
In principle, error-success revisionists could be either sort of success theory— libertarian or compatibilist. Though it is a conceptual possibility, libertarian moderate revisionism will strike many as wrong-headed. First, one would have to develop an account of the error that did not make some libertarian element its source. Second, acceptance of a folk conceptual error theory, even when coupled with a success theory about the property of responsibility, would undermine a key motivation for libertarianism. If one thought the folk concept was both libertarian and error ridden, why not opt for a revision that does not require a defense of libertarianism and all the difficulties that entails? Why hold on to the libertarianism at all? Error-success libertarians (to the extent to which there are or will be any) will reply that since our concern is with responsibility and the kind of agency it requires, if responsibility requires libertarianism we should be revisionist libertarians.[27]

The path to error-success libertarian revisionism is not so easy, though. Most existing arguments for incompatibilism rely on the intuitive plausibility of principles whose main recommendation seems to be that it best captures ordinary thinking. But, as we saw earlier, we

cannot just blindly suppose our ordinary concept accurately reflects the facts about responsibility. Some further argument is needed to show that standard incompatibilist principles reflect facts about the world beyond the considered intuitions of highly educated people in the English-speaking world. Supposing that some such account were offered, the libertarian would then face a question of significance: why care about responsibility if it is not like we imagined? The most likely answer will point to the normative significance of responsibility— that is, we ought to care about responsibility because its demands are somehow binding or otherwise relevant to us. But, unless we can clearly show that the normative bindingness of responsibility practices (and not just our folk concept) *required* libertarian agency, there is considerable pressure to adopt an alternative revisionism that (1) tracks the normative features we are concerned to capture and (2) does not require the demanding metaphysics of libertarianism. Such is the project of the other main form of moderate revisionism.

The other success theory incarnation of moderate revisionism is compatibilist, though compatibilist in a way distinct from weak revisionist and non-revisionist compatibilism. The considered compatibilism maintains that our folk concept of responsibility really is implausible or error-ridden (which the weak revisionist and non-revisionist cannot admit), but that despite these acknowledged errors, there is some property of responsibility to which we can refer.

Given these two possibilities, we can finish our revision of R to better reflect the most viable strands of revisionist theorizing. Consider R* (fig. 4):



My discussion and R* are concerned with the main strands of revisionism, though there are other, less appealing, possibilities (e.g., weak revisionist libertarianism and non-revisionist error-error theories). I take it, though, that I have provided the conceptual apparatus to map those views as well, though I will not do so here.[28]

6. The contours of moderate revisionism

Though few have done so, there are good reasons to specify when a theory is moderately revisionist or not. The main reason concerns the distinct burdens and benefits of moderately revisionist theories. Moderate revisionist theories start with the advantage of not having to deny the plausibility of arguments for incompatibilism when they are construed as arguments about our folk concept of responsibility. In the case of the compatibilist moderate revisionist, this means that standard objections about compatibilism's failure to capture the full, intuitive notion

of responsible agency have less force. Our failure to recognize the moderately revisionist elements in a theory — which is just as often a failure of self-recognition — potentially means that significant amounts of intellectual resources have been squandered responding to objections that do not apply to a moderately revisionist theory. Once we are clear about the revisionism, we can have a theory that can accept what is most persuasive about incompatibilism without giving up the more naturalistically plausible metaphysics of compatibilism.

The burdens of moderate revisionism ought not go unnoticed as well. Since the error-success theorist thinks that there is a folk conceptual error and that it does not infect reference, the error-success theorist needs to say both what the conceptual error is, why it does not infect reference, what the property of responsibility amounts to, and how we can know anything about it, given a systematic error in our thinking about it. To date, few accounts that might be read as moderately revisionist have taken any steps to answer these questions.

Who, then, is a moderate revisionist? The possibility of inadvertent or unacknowledged revisionism makes identification of moderate revisionist theories a tricky business.[29] For example, as the O'Connor quote at the start of the paper suggests, some might be inclined to read event-causal libertarian views such as Kane's or Ekstrom's as a kind of moderate revisionist libertarianism.[30] At least Kane rejects this interpretation of his view and Ekstrom seems likely to do so as well. And, given what we have seen of the burdens and difficulties facing specifically libertarian moderate revisionist theories, it does not seem an especially fruitful route for them to pursue. Thus, if they are paradigmatic revisionists— which I doubt— they are most likely revisionists of an inadvertent or non-paradigmatic sort.[31] In any event, it remains an open question whether or not there are any genuine cases of libertarian moderate revisionism.

As suggested in the previous section, compatibilist moderate revisionism seems more promising. And, a group of compatibilists influenced by P.F. Strawson might, with some charity, be read as at least proto-moderate revisionists. Philosophers who focus on "the varieties of free will worth wanting" or argue for limited revisions of retributivist elements in responsibility practices— e.g., Dennett, Wallace, Scanlon, and so on— might all be interpreted as moderately revisionist in different ways.[32] That is, we might understand these theories as attempting to revise our ordinary concept of responsibility in such a way that it better reflects an accurate understanding of the property of responsibility. Still, the amount of charity involved in interpreting these accounts as moderate revisionist theories, whether inadvertent or not, suggests that such an interpretation would be as much recasting as interpretation.

J.J.C. Smart's under-appreciated revisionist account in "Free Will, Praise, and Blame" is an interesting candidate because aspects of his view demand considerable revision away from ordinary practices.[33] Smart is explicitly revisionist about free will, praise, and blame. He calls for the replacement of blame with something he calls "dispraise," because he holds that blame presupposes implausible libertarian metaphysics. Dispraise, however, only involves a form of "grading" or evaluation that does not presuppose an implausible picture of agency. Notably, he treats the status of praise and blame as distinct from the status of responsibility, for he believes that responsibility amounts to susceptibility to moral influence, and that such susceptibility is not affected by troubles with libertarian metaphysics. Whether his proposed elimination of blame practices can be described as strong revisionism about the practices characteristic of responsibility, and whether that entails moderate or strong revisionism about the concept of responsibility, depends on several issues: (1) the presumed semantic theory for responsibility, (2) the degree to which responsibility has a genuine place in utilitarianism, and (3) whether praise

and blame can be separated from a theory of moral responsibility.[34] These raise a number of thorny issues I will not pursue here, but this suggests that borderline cases such as Smart's will require considerable unpacking to classify with precision.

The complexity of unpacking Smart's case is instructive, and suggests that explicitly revisionist proposal will need to specify the sense(s) in which prescriptive proposals are intended to be revisions. Consider other revisionist theories, including those recently proposed by Pereboom, Honderich, and Smilansky.[35] They hold that there is a very important sense in which we are not responsible, and that if we were to be entirely accurate about things (though Smilansky does not think we ought to be), we would expunge from our language and practices those elements that presume that notion of responsibility. To that extent, they will count as strong revisionists.[36] However, they each also argue that the integrity of a large subset of our responsibility-characteristic practices and attitudes remain unaffected by determinism. Depending on what we think about those unaffected practices and attitudes, we might go on to hold (as they do not) that we are indeed fully responsible, but that responsibility turns out to be a bit different than we might have imagined. Thus, though Pereboom, Smilansky, and Honderich do not think of themselves as committed to an error-success theory of this sort (because they insist that we are not responsible in some deep sense— and, at least in the case of Honderich his non-cognitivism precludes this possibility), there is no obvious reason why a moderate revisionist could not accept much of what they say and simply insist that what this shows is that the property of responsibility is merely different than they supposed. In sum, though there are some accounts that may be inadvertently moderate revisionist, proto-revisionist, or well-suited to recasting as moderate revisionist, the landscape of possibilities has hardly been capitalized upon.

Still, reflecting on the landscape can generate useful insights for existing projects. For example, moderate and strong revisionists could agree that a sizable subset of our responsibility-characteristic practices is in good standing. If so, this points to considerable room for shared projects between two classes of theory that ordinary categories treat as polar opposites: (error-success) compatibilism and (error-error) hard incompatibilism.[37] By revisionist lights, the best versions of compatibilism and hard incompatibilism can be understood to have a shared project of settling whether the subset of determinism-immune practices is enough for responsibility, whether those practices and attitudes should be changed in some way, and what other changes follow as a result of abandoning parts of the framework of ordinary thinking about responsibility. The traditional compatibilism/ incompatibilism framework makes this insight difficult to see, but a revisionist framework helps reveal it.

Reflecting on the way traditional categories obscure widely shared projects in the theory of moral responsibility can help us to acknowledge an important point— we have paid entirely too much attention to the labels of traditional philosophical categories. At the end of the day, what we need to know from a theory of responsibility is how we are to treat one another in the relevant contexts. Once we clarify what practices, attitudes, and beliefs are normatively sound irrespective of our folk beliefs, whether we call such a theory compatibilist or incompatibilist may be a matter of semantics in the colloquial sense— a distinction without much difference. You might never learn this, though, unless you thought about traditional categories in light of revisionist possibilities. What this shows is that we should think about responsibility in revisionist-friendly terms so as to speed along the dismantling of our entrenched, but ultimately disadvantageous current philosophical categories. Much of the urgency concerning debates about the compatibility of determinism and our commonsense concept of moral responsibility

might dissolve if we recognize how widely shared the project is of justifying, refining, and revising our responsibility characteristic practices, attitudes, and concepts. Of course, there will be disagreements on what and how much revision is required, and some of this will still turn on issues of compatibility. However, once we see that there is be no prima facie problem with being a suitably sophisticated deflationist, there is less incentive to worry whether proposed theories fully capture ordinary thinking, and thus, whether ordinary thinking is really committed to traditional forms of incompatibilism or not.

7. Making it explicit

We have seen that there are several things that might recognizably count as revisionism. The central case, paradigmatic revisionism, is a theory that prescribes revision, relative to a diagnostic account of commonsense thinking, in our responsibility-characteristic practices, attitudes, beliefs, or our conception of these things. Since the taxonomy of paradigmatic revisionism I have offered is neutral with respect to the issue of *de re* revisionism and the semantics of responsibility, it offers a way to classify virtually all theories as paradigmatically revisionist or not, without presupposing that one account of responsibility has correctly characterized the nature of responsibility or our concept of it. Since inadvertent revisionism is a possibility, and since philosophers thus far have been largely inattentive or insensitive to the implications of revisionism, we have seen how classifying existing accounts as revisionist or not raises some interpretive challenges. This should be a lesson for both critics and proponents of revisionism: objections to revisionism will need to be more fine-grained than they have been thus far and theories undertaken in a revisionist spirit will need to me much clearer about their revisionism than they have been.

Revisionism, especially of the moderate sort, has some attractions, not the least of which is that it has a way of balancing some traditional incompatibilist and traditional compatibilist concerns, the former in its diagnostic account and the latter in its prescriptive account. Perhaps a more important upshot of thinking in broadly revisionist terms, however, is that it both brings to our attention a largely neglected space of theoretical possibilities while at the same time explaining what unifies seemingly diverse projects such as compatibilism and hard incompatibilism.

Revisionism is not without its own challenges, however. For example, one might expect that critics of revisionism will argue that explicitly moderate revisionist theories are ersatz theories of responsibility, or that they are in some other way inadequate theories. Though more precise responses will have to await more precise criticisms, a moderate revisionist response is already latent in the remarks I have made. Even if turns out that, a given moderate revisionist theory is not a theory of responsibility, strictly speaking, but a theory of responsibility*, it will nonetheless have a right to our attention because it will (1) capture much of what we were concerned with when we set about constructing a theory of responsibility and (2) it will be a theory with justified normative force. And, if this is true, we might then wonder why we should care about responsibility as opposed to responsibility*, unless the theory of responsibility also satisfied the plausibility and normative adequacy conditions. And, if it did, that too would be an interesting result worth more reflection for we have overlapping justified normative systems that entail conflicting answers to particular cases. Thus, the revisionist will argue that we have good reason to forge ahead, even if we are skeptical, uncertain, or agnostic about the semantical and metaphysical issues that would decide whether the theory is, strictly speaking, a theory of responsibility. In sum, independent of how the substantive metaphysical and semantical issues

sort out— including debates about the compatibility of determinism with our commonsense concepts or the thing they refer to— there is reason to desire a theory of the sort that moderate revisionism endeavors to offer: one that can guide our responsibility-characteristic practices, attitudes, and beliefs without presupposing an implausible picture of agency, psychology, or society.

If the account I have offered is correct, revisionism in several forms is already a pervasive part of our theorizing about responsibility. If we are to take full advantage of the moderate form of paradigmatic revisionism, however, we cannot leave it implicit, accept it as a last resort, or inconsistently endorse it. This is true for both sides of the conventional compatibility debate. The accidental revisionism of some compatibilists needs to give way to a more thorough reflection on the significance of revisionism for compatibilism. Though doubtlessly some features would be changed, most forms of compatibilism are likely to find new resources when donning revisionist garb. For their part, incompatibilists have much to gain by developing more sophisticated pictures of what options are available to us besides libertarianism and complete eliminativism. In particular, error-error theories are likely to find new plausibility if they are supplemented with a combination of conservatism and moderate and strong revisions.

While revisionists will differ about whether we really are responsible, they are united in focusing on the question of how it is that we ought to go about revising our collection of beliefs, attitudes, and practices characteristic of responsibility. Recall the morning of the fabled declaration of universal determinism. On this morning, moderate revisionist of many stripes will smile, perhaps stretch, and begin propagating a theory of responsibility that respects the psychology we have, the world we live in, and the norms that make our attitudes and practices justifiable. What more should we expect from a theory of responsibility?[38]

[1] Daniel Dennett *Elbow Room* (Cambridge: MIT, 1984). Other works that sometimes give expression to this theme include Susan Wolf's *Freedom Within Reason* (Oxford: Oxford, 1990), Jonathan Bennett's interpretation of P.F. Strawson in "Accountability" in Zak van Straaten, ed., *Philosophical Subjects* (New York: Clarendon, 1980), and at least with respect to revisionism about retributivism, R. Jay Wallace in *Responsibility and the Moral Sentiments* (Cambridge, Harvard, 1996) and T. M. Scanlon in both "The Significance of Choice" in *The Tanner Lectures on Human Values* VII, ed. by McMurrin (Cambridge: Cambridge, 1988) and *What We Owe to Each Other* (Cambridge: Belknap Harvard, 1998). For an evaluation of the broadly Strawsonian approach to revisionism, which includes many of the figures just mentioned, see my "Responsibility and the Aims of Theory: Strawson and Revisionism" *Pacific Philosophical Quarterly* (forthcoming).

[2] Today there are few (if any) hard determinists of the traditional sort. Important successor views to hard determinism include Derk Pereboom *Living Without Free Will* (New York: Cambridge, 2001) and Galen Strawson "The Impossibility of Moral Responsibility" *Philosophical Studies* 75 (1994) and "The Bounds of Freedom" *The Oxford Handbook on Free Will*, ed. Robert Kane (New York: Oxford, 2002), pp. 441-460. See also Ted Honderich *A Theory of Determinism* (New York: Oxford, 1988), and Saul Smilansky *Free Will and Illusion* (New York: Oxford, 2000) for theories that there is some sense in which we are responsible and some sense in which we are not.

[3] Fischer and Ravizza hold that free will is not compatible with determinism, but that moral responsibility is—hence the name "semi-compatibilism." See Fischer and Ravizza, *Responsibility and Control* (Cambridge: Cambridge, 1998).

[4] For example, see McGinn, *Problems in Philosophy* (New York: Blackwell 1994), van Inwagen, Peter, "Review of *Problems in philosophy*," *The Philosophical Review.* (1996) 105, 2: 253-256, and Timothy O'Connor, *Persons and Causes* (New York: Oxford, 2000), p. xii. McGinn's rejection of revisionism (understood as "deflationism" in his language) is tied to his view that traditional philosophical problems and their purported solutions are manifestations of a widespread human cognitive shortcoming.

[5] Op. cit., pp. 41-2.

[6] McGinn, op. cit., and van Inwagen, op. cit., 253-256.

[7] *Persons and Causes*, p. xii.

[8] For instance, R.J. Wallace in *Responsibility and the Moral Sentiments*, p. 288 tentatively allows for a "moderately revisionist" interpretation of his theory of responsibility and desert. Similarly, in *What We Owe To Each Other*, p. 294, Scanlon concedes that he would be content if his account of responsibility turns out to be revisionist with respect to retributivism, though he does not build the theory with revisionist aspirations. In *Elbow Room*, op. cit., Dennett moves back and forth between defending a theory of our commonsense concepts and advocating some revisionism. Even if Kane concedes that his theory is revisionist in the way O'Connor claims, revisionism is certainly not an intended or central part of the main statement of his theory in *The Significance of Free Will* (New York: Oxford, 1996), a point he has emphasized in personal correspondence.

[9] A possible example of unconventional but principled *de re* revisionism would be a theory that attempts, via its prescriptive aspect, to re-anchor our practices and attitudes in a different set of properties, but in doing so, making that new set of properties the properties of responsibility. The idea would be to change the nature of responsibility by changing the way we think about it, and/or the way we engage in its characteristic practices and attitudes. If this is not possible, then it is much harder to see how intended and principled *de re* revisionism is possible. Less contentiously, we can allow for the possibility of inadvertent *de re* non-revisionism. A theory that offers a prescriptive proposal that differs from its diagnostic proposal (the central variety of revisionism with which we are concerned), but where the diagnosis is wrong and the prescription turns out (in spite of its intended revisionism) to describe the nature of responsibility would be an instance of inadvertent *de re* non-revisionism.

[10] The gap between our self-construals and our responsibility characteristic practices, attitudes, and beliefs is explored in greater detail in section 4.

[11] Revisionism of the *de re* sort does not, by itself, entail revisionism of the *de facto* sort, and vice-versa, though some directions of entailment might follow given one or another semantics for 'responsibility'. Excluding this latter possibility, the antecedent point can be easily illustrated. A *de re* non-revisionist theory might be a *de facto* revisionist theory if it prescribes changes in our practices. This could occur when we set about correcting an inconsistency in our practices in light of The True Theory of Responsibility. A *de re* non-revisionist theory can also be *de facto* non-revisionist. Something like this seems to be the hope lurking behind conventional theories of responsibility, that we might have a theory that accurately describes the nature of responsibility without requiring changes in our responsibility-characteristic practices or thinking. A *de facto* revisionist theory might also be coupled with a *de re* revisionist theory. This possibility could be instantiated by a theory that misdescribes the nature of responsibility in conjunction with prescriptions for changes in our responsibility practices or thinking. And, of

course, a *de facto* revisionist theory that is *de re* non-revisionist is merely the case I described first, with the order of description changed.

[12] See Geoffrey Sayre-McCord, "The Many Moral Realisms" in Sayre-McCord, ed. *Essays on Moral Realism* (Ithica, Cornell, 1988).

[13] Richard Double, *The Non-Reality of Free Will* (New York: Oxford, 1991), and Pereboom *Living Without Free Will*, pp. 131-133.

[14] Honderich, op. cit., and P.F. Strawson "Freedom and Resentment" *Proceedings of the British Academy* 48 (1962), pp. 1-25. I discuss some difficulties with Strawson's non-cognitivism in "Responsibility and the Aims of Theory" *Pacific Philosophical Quarterly* (forthcoming). Important discussions of Strawson's non-cognitivism include Susan Wolf "The Importance of Free Will" *Mind* 90 (1981), pp. 386-405 and Gary Watson "Responsibility and the Limits of Evil" in Ferdinand Schoeman, ed., *Responsibility, Character and the Emotions* (Ithica: Cornell 1987).

[15] There are a variety of ways to spell out the truth conditions under cognitivist construals of responsibility, and not all of them are objectivist (where objectivist means something like mind- and convention-independent). Further, there is no *prima facie* reason why a worked-out theory of responsibility could not involve a variety of different truth conditions. For example, one might propose a subjectivist epistemic condition on responsibility while maintaining an objectivist interpretation of the freedom condition. One might also split up the satisfaction of a single condition over subjective, intersubjective, and objective features of the world.

[16] This is not always true, however. Consider that many compatibilists argue against libertarianism on the grounds that indeterminism disrupts responsibility. For these compatibilists, if it turned out that the world had indeterminacies of just the right sort (or perhaps more accurately the 'wrong sort'), then we would lack responsibility under even compatibilist understandings of responsibility. Though a complete classification system would represent this possibility as a species of error-theory, I do not know of anyone who has developed such a theory, though for a broader discussion of these issues see Ishtiyaque Haji "Alternative Possibilities, Luck, and Moral Responsibility" *The Journal of Ethics* 7.3 (2003), pp. 253-275.

[17] Though I think the description of the conditionalist position is (as a general characterization) largely true, it is open to the conditionalist to maintain that people really do understand the relevant notion of 'can' in a conditional way. If so, then the conditional analysis is not even a form of weak revisionism and it is instead merely an accurate reporting of what we believe and what is true (or so they must maintain). However, if undergraduates are any indication, it is implausible to maintain that the folk's responsibility-relevant notion of 'can' is typically understood

by the folk as committed to anything like standard conditional analyses. Of course, the folk could simply misunderstand their own responsibility-relevant notion of 'can'. However, that just brings us back to the description of conditionalists as weak revisionists.

[18] Discussing the intricacies of these cases and their varied characterizations would take us too far a field. However, many important essays on this issue, including Frankfurt's original, are collected in David Widerker and Michael McKenna, eds. *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (Aldershot, England: Ashgate, 2003). Also, see Fischer's helpful discussion in J.M. Fischer "Recent Work on Moral Responsibility" *Ethics,* 110 (1999), pp. 93-139.

[19] See Fischer and Ravizza, op. cit., and various other places, including Fischer "Responsibility, History, and Manipulation" *Journal of Ethics*, 4.4 (2000), especially in a response of his to Kane on pp.385-6.

[20] See citations in n.2. This list includes Smilanksy and Honderich, though they would such inclusion. Nonetheless, I believe it is an apt one. They endorse something we can call *The Honderlansky Thesis*. The Honderlansky Thesis is that free will and/or moral responsibility is both compatible and incompatible with determinism. The idea is that there are different senses of freedom or responsibility, and that some are compatible with determinism and some are not. Be that as it may, The Honderlanksy Thesis encourages unhelpful departures from standard usages of 'compatibilism' and 'incompatibilism'. I take it that what is at stake in debates moral responsibility and free agency is whether free will and moral responsibility, *understood in the fullest sense*, is compatible with determinism (or some other picture of the world). Many incompatibilists (e.g., Kane, G. Strawson) admit that there are senses of freedom and even responsibility that are compatible with determinism. What is at stake is whether the most robust widely shared notion we have is compatible or not. Since Smilansky and Honderich both admit that the strongest sense of responsibility is *not* compatible with our most plausible pictures of the world, for my purposes (and, I suspect, for the purposes of most philosophers who work on free will and moral responsibility) they are effectively hard incompatibilists. For a similar reaction, see Pereboom, *Living Without Free Will*, p. 131.

[21] Smilansky (in my framework, a hard incompatibilist— see the previous note) resists eliminativism on interesting pragmatic grounds— he thinks we are better off living with the illusion of free will and moral responsibility than we would be if we widely recognized that purportedly illusory nature of these things.

[22] Given my prior claim that conditional analysis compatibilists are typically weak revisionists, and hard incompatibilists are typically strong revisionists, one might wonder who are the moderate revisionists. A more complete answer has to await the introduction of several additional distinctions, though a provisional answer can be

given: some Strawsonians— though perhaps not intentionally, and certainly not explicitly or exclusively— and perhaps a handful of other figures.

[23] Depending on how you think about properties, you can either characterize a traditional error theory as one in which there is no property of responsibility or one in which the property of responsibility is not instantiated. This difference is inconsequential for our purposes. I should also note that in his "The Mad Scientist Meets the Robot Cats" *Philosophy and Phenomenological Research* (1996), pp. 333-337, Mark Heller has argued for the importance of distinguishing between the concept of free action and free action itself. I am agnostic about the semantics of responsibility.

[24] See John M. Fischer and Mark Ravizza, *Responsibility and Control*, op. cit., for this particular example, though P.F. Strawson argues along similar lines in "Freedom and Resentment."

[25] For an example of this, see Saul Smilanksy "The Ethical Advantages of Hard Determinism" *Philosophy and Phenomenological Research* 54 (1994), pp. 355-366. Because theorists have not always been clear on the differences and because it oftentimes depends on the semantics of revisionist metaphysics, it is frequently difficult to make out whether or not a given proposal is intended in a moderately or strongly revisionist vein. So, for closely related projects, also see Lawrence Stern "Freedom, Blame, and the Moral Community" *Journal of Philosophy* (1974), pp. 72-84, and Saul Smilansky *Free Will and Illusion* (New York: Oxford, 2000).

[26] This is an objection that many hard determinists have faced. For moderate revisionists, it merely represents a likely constraint on revisions.

[27] I discuss whether event-causal libertarians such as Kane (*The Significance of Free Will*, op. cit.) and Laura Ekstrom (*Free Will: A Philosophical Study*, Boulder, Colorado: Westview, 2000) are libertarian moderate revisionists in § 6.

[28] It is also worth noting that by focusing on conceptual revision, there are kinds of revisionist theories that R* does not capture. For example, I have not accounted for revisionist theories that do not think our concept needs any sort of revision, but that some of our attitudes do need revision.

[29] This excludes, I hope, my own recent developments of this approach in "Responsibility and the Aims of Theory," "History and Responsibility" (under review), and "Moral Influence, Moral Responsibility" (under review).

[30] Robert Kane, *The Significance of Free Will* (New York: Oxford), 1996 and Laura Waddell Ekstrom, *Free Will: A Philosophical Study* (Boulder, Colorado: Westview), 2000.

[31] If we want to make sense of O'Connor's claim, we either have to interpret it as claiming that Kane is a *de re* revisionist, a *de facto* revisionist, or an inadvertent moderate revisionist. It is not particularly charitable to interpret O'Connor's remark as an attribution of *de re* revisionism because as an agent causalist, he likely believes that all views that are not full-bloodedly agent causal are revisionist in the *de re* sense. This would make his remarks about Kane trivial. Alternately, he could be suggesting that Kane's view is a departure from the facts about how we think about responsibility, independent of the nature of responsibility itself. This interpretation has a certain degree of plausibility about it. However, given O'Connor's apparent acceptance of the unity of diagnostic and prescriptive projects, this interpretation amounts to little more than the prior objection. This leaves the interpretation that makes Kane out to be an inadvertent moderate revisionist. To make *this* charge stick, O'Connor would have to show a difference between the diagnostic and prescriptive aspects of Kane's account. It is not obvious that he has done this, though perhaps he thinks it could be done.

[32] I explore this possibility at greater length in "Responsibility and the Aims of Theory."

[33] *Mind* 70 (1961), pp. 291-306.

[34] If Smart is right that revisionism about free will, praise, and blame does not entail revisionism about responsibility, then he is an early semi-compatibilist, i.e., one who believes that free will is incompatible with determinism, but that responsibility (moral influence) is compatible. On most contemporary accounts of responsibility, though, free will, praise, and/or blame *are* connected to responsibility in a way that makes revision of at least some of these things constituitive of revisionism about responsibility. It is also worth noting that the view of moral responsibility as moral influence has been widely rejected. For examples, see P.F. Strawson, op. cit., Jonathan Bennett, "Accountability" in Zak van Straaten, ed., *Philosophical Subjects* (New York: Clarendon, 1980); and T.M. Scanlon "The Significance of Choice" in S. M. McMurrin, ed. *The Tanner Lectures on Human Values* VIII (Cambridge: Cambridge, 1988). For an attempted rehabilitation of this view, see Vargas "Moral Influence, Moral Responsibility."

[35] Smilansky, *Free Will and Illusion*, and Pereboom, *Living Without Free Will.*

[36] In the case of Smilansky, some qualification is in order. He holds the view that free will is an illusion we ought not forfeit (a view he calls "illusionism"). Depending on the degree to which his theory, by its own lights, requires revision of our practices, attitudes, or beliefs, his might be construed as a sophisticated version of a *non*-revisionist error-error theory.

[37] John Martin Fischer suggests something similar in "Excerpts from John Martin Fischer's Discussion With Members of the Audience" *Journal of Ethics* 4.4 (2000), p.415. There, he reports that he and Pereboom agree that their views "are actually very close." Though it would require recasting of Fischer's view for it to count as an error-success version of compatibilism (it would, I think, be a stretch to treat his weak revisionism about the concept of responsibility as a folk conceptual error theory about responsibility) I take it that my account explains why such an agreement need not be just the possible effects of wine, as Fischer wryly suggests in the previously cited passage.

[38] This paper and its ancestors have benefited from the comments of many people, including: Michael Bratman, John Martin Fischer, Nadeem Hussain, Tim Schroeder, Saul Smilansky, Ken Taylor, Derk Pereboom, R. Jay Wallace, the terrific crew of philosophers at Caltech (especially Alan Hájek), the participants in my Free Will seminar in the Spring 2003, and the participants at my session in the 2002 Eastern APA meeting. I also want to thank an anonymous reviewer at *Philosophical Studies* for feedback that made this paper considerably better, as well as two prior anonymous reviewers who had the misfortune of reading an earlier and less developed version of this paper. I am also grateful for the financial support I received from the USF Fleishhacker Family Endowment and the California Institute of Technology while working on parts of this paper.