RESPONSIBILITY AND THE AIMS OF THEORY: STRAWSON AND REVISIONISM

BY

MANUEL VARGAS

Abstract: Strawsonian approaches to responsibility, including more recent accounts such as Dennett's and Wallace's, face a number of important objections. However, Strawsonian theories can be recast along revisionist lines so as to avoid many of these problems. In this paper, I explain the revisionist approach to moral responsibility, discuss the concessions it makes to incompatibilism (including the point that compatibilists may not fully capture what our commonsense understanding of responsibility), why it provides a fruitful recasting of Strawsonian approaches, and how it offers an alternative to the pattern of dialectical stalemates exhibited by standard approaches to free will and determinism.

In recent years, reflection on the relationship between individual moral responsibility and determinism has undergone a remarkable renaissance. Incompatibilists, those who believe moral responsibility is incompatible with determinism, have offered powerful new arguments in support of their views. Compatibilists, those who think moral responsibility is compatible with determinism, have responded with ingenious counterexamples and alternative accounts of responsibility.

Despite the admirable elevation of complexity and subtlety within both camps, the trajectory of the literature is somewhat discouraging. Every dialectical stalemate between incompatibilists and compatibilists seems to be superseded by a similar though often more subtle stalemate.[1] The stalemates have two sources. On the one hand, incompatibilists again and again find powerful intuitive support from our folk concept. On the other hand, compatibilists seem right to insist that even if determinism were true, this would not mitigate our need for a concept of responsibility.

Pacific Philosophical Quarterly 85 (2004) 218–241

In this paper, I attempt to show how principled and systematic pursuit of an approach I call *revisionism* might push us through this stalemate. The central idea of revisionism is that an adequate theory of responsibility will depart significantly from our commonsense understanding of responsibility. My point of departure is P. F. Strawson's justly influential "Freedom and Resentment" and some of the work that article has inspired.[2] I start with Strawsonianism because careful attention to *how* it fails suggests a way to rehabilitate it along systematically revisionist lines. This recasting requires that we make some important concessions to incompatibilists, including the idea that no compatibilist theory may be able to respect the constraints of the ordinary concept of moral responsibility.[3] However, if we adopt my approach we have good reason to think we can make real progress against the pattern of dialectical stalemates.

This paper is divided into three parts. In part one, I describe Strawson's original account and the main lines of criticism it provoked. In part two, I argue that no standardly compatibilist Strawsonian account has the resources to answer traditional incompatibilist worries. In part three, I say what the revisionist alternative is, how it can be pursued, and why it constitutes a promising alternative to standard forms of compatibilism and incompatibilism.

1.

A. STRAWSON ON RESPONSIBILITY

In "Freedom and Resentment," Strawson sought to put compatibilism on a new and more persuasive footing. Strawson aimed to give the compatibilist an account of our moral practices showing that they were justified and did not depend on, as he memorably put it, the "panicky metaphysics" of libertarianism. He argued that responsibility is to be understood in terms of a set of distinctive attitudes and associated practices. According to Strawson, if we carefully reflect on the way these attitudes and practices function, we would find nothing internal to our practices to suggest that the truth of determinism should prevent us from engaging in those practices.

Strawson described two kinds of cases in which we do not hold people responsible: (1) when agents act in a way that does not reflect a poor "quality of will" and (2) in cases where the target of assessment is not the right sort of agent to be a target of our practices and attitudes. I will follow Watson in identifying the suspension of the attitudes characteristic of holding people responsible in the former case as *excuses*, and in the latter case as *exemptions*.[4] Excuses do not depend on truths about determinism. When we excuse someone, we typically do so because the targeted agent fails to have a criticizable quality of will. For example, if I accidentally

step on your foot, I will be excused from responsibility not because determinism entailed that I stepped on your foot, but rather because I did not step on your foot out of ill will. According to Strawson, my ill will is what matters for our attitudes and practices, not whether my will had been determined. The case of exemptions is similar: when we exempt agents from responsibility, it is not because they are determined, but because they are simply agents of the wrong kind.[5] These agents typically lack the right sensitivity to moral practices because they have not yet developed the relevant capacities (children), they have lost them (the injured, diseased, or aged) or the agents never had them. It is a departure from ordinary adult capacities that exempts agents from responsibility, not the threat of determinism. In short, neither excuses nor exemptions are sensitive to abstract truths about determinism.

Since the truth of determinism does not affect our practices through either excuses or exemptions, nothing internal to the practices of holding people responsible suggests that we need to be worried about determinism. But, Strawson recognized that one might challenge the entire framework of practices and the attitudes they express. The critic might argue that the framework presupposes something false or that it stands in need of some further justification. Strawson responded that the practices (and the attitudes they express) are part of an inescapable framework of interpersonal relationships and thus do not require further justification.[6] There is a strand in Strawson's discussions that suggests that questions about this framework are unintelligible, for the framework in question is somehow foundational or necessary for our thinking about responsibility. The main thrust of his argument, though, is that because we cannot help having attitudes that give rise to our practices of holding people responsible, demand for further justification is inappropriate. The demand for justification of our practices comes to an end when the root of those practices, the responsibility-characteristic attitudes, turns out to be inescapable features of human psychology. Finally, Strawson suggests a possible pragmatic defense. Even if we could give up the framework of attitudes, this would be to give up many of the rich social-psychological features that make our lives worth living. Consequently, determinism does not and could not pose a real threat to responsibility.

B. REACTIONS TO STRAWSON

Strawson's theory provoked criticism from a number of directions. Several philosophers have offered rigorous criticisms of various aspects of Strawson's substantive account of our practices and what they require. For our purposes, we need not consider them in any detail, though they point to something widely recognized about Strawson's account: his analysis of our practices, while insightful and suggestive, is seriously underdeveloped.[7]

In particular, Strawson says too little about who is exempt from responsibility and why. One way of understanding the threat of determinism is to think that it might show that we are all exempt from responsibility because we are not the kinds of agents whom it would be appropriate to treat as responsible in the sense presupposed by our practices.

A second line of criticism concerns the claim that our responsibility practices and attitudes are an inescapable part of the basic framework of human social life.[8] Several philosophers question whether the attitudes characteristic of responsibility – what Strawson calls the *reactive attitudes* in cases in which one is reacting to some personally directed responsibility-bearing act, and the *vicarious analogs* in cases in which one is responding to an action directed at others – are truly inescapable human reactions.[9] If these critics are right, then the responsibility-characteristic attitudes *can* be called into question precisely because they are malleable in a way that does not presuppose abandonment of our entire network of interpersonal attitudes. Consequently, the justification of our responsibility-characteristic attitudes, and the practices that depend on them, is still in order.

Although it is far from clear that our attitudes are as plastic as these critics of Strawson suggest, others have convincingly argued that we can intelligibly raise questions about the framework of attitudes even if the attitudes are ultimately inescapable. This third sort of criticism can (though it need not) concede the inflexibility of our psychology, but resists Strawson's claim that these attitudes cannot be subject to further justificatory demands because of their place in the framework of our lives. It could turn out, Susan Wolf argues, that we are not truly responsible even though we cannot help treating people as though they are responsible agents.[10] Believing people are responsible and treating agents as responsible would be species-wide instances of what Dennett calls "the familiar class of life-enabling or life-enhancing illusions: the illusion that one is still loved by one's loved ones; the illusion that one has several more years to live when one hasn't; the illusion that in spite of one's physical ugliness, one's inner beauty is readily manifest to others."[11] Even if our responsibility-characteristic attitudes are inescapable, there are other parts of our doxastic and value framework from which questions about our practices might be raised. To paraphrase Wolf, our interest in living in accord with the truth can ground challenges to even highly implastic attitudes. On this line of criticism, the framework of responsibility turns out to be only a sub-set of our more complete conceptual, axiological, and connative framework. It is from the perspective of other aspects of that framework that the responsibility framework can be called into doubt.

One way to get at this family of criticisms is to focus on the apparently cognitivist character of responsibility. This is the idea that claims of responsibility admit of truth and falsity, and our "grounding beliefs" – the beliefs that provide the foundations for our judgments that someone

is responsible – might well be false. Call this worry the *cognitivist criticism.* To see why we might worry about the apparently cognitivist dimension of responsibility (i.e., the grounding beliefs), we need only recall the hackneyed example of judgments about the location of the sun in pre-Copernican times. Before the Copernican revolution changed much of our thinking about cosmology, people believed that during midday the sun was objectively above them. This judgment was made against a background of other beliefs, including fixed and absolute spatial relations. The problem was, of course, that some of these background beliefs were false. The worry that motivates the question of responsibility is similar: Do our practices and attitudes characteristic of responsibility depend on a judgment that presupposes false or incoherent things? Strawson downplayed this worry, instead directing our attention to the inescapability of the reactive framework. But Wolf and others are right to insist that we cannot rule out this question and the metaphysics that its answer might bring, simply by declaring that the attitudes triggered by our judgments are an inescapable part of our social psychology. Assuming that we do care about whether or not we are truly responsible in the ordinary sense of the phrase, Strawson needs to show that our expression of our attitudes does not depend on a judgment or judgments (however inevitable) that is or are false. Strawson never does this.

Bernard Berofsky recently noted "Strawson's celebrated proposal to construe freedom and responsibility as constituitive of human society failed to convince enough of us that metaphysical issues cannot have a bearing on the attitudes and perhaps even the practices associated with these notions."[12] Berofsky's comment points to an important similarity between two of the three lines of criticism I have mentioned. The point of the first criticism was that Strawson lacks a sufficiently detailed account of when someone is or is not exempt from responsibility. The upshot of the third criticism, the cognitivist criticism, is that Strawson lacks a sufficiently detailed account of our grounding beliefs for responsibility ascriptions. In both cases, the main worry turns out to be the possibility that a more complete account of the conditions for responsibility will invoke a troublesome metaphysics, which might undermine the normative integrity of the existing practices. In short, Strawson needs to be given something more to say.

2.

A. STRAWSONIANS

Several contemporary compatibilists have attempted to rehabilitate Strawson's theory in a way that answers the cognitivist criticism and the

more general implication that a full accounting of our beliefs might or must yield an unacceptable metaphysics. Among the most sophisticated and thorough attempts to do so are R. Jay Wallace's *Responsibility and the Moral Sentiments* and Daniel Dennett's *Elbow Room.*[13] Both works attempt to show that the beliefs presupposed by judgments of responsibility are neither metaphysically robust nor especially troublesome. Here, I will show that even sophisticated Strawsonian theories such as these cannot give an adequate accounting of the beliefs that matter for our judgments of responsibility. These theories answer the cognitivist criticism because they can accommodate the cognitive structure of responsibility claims and beliefs. Given the kinds of things these theories postulate in accounting for those cognitivist features, however, incompatibilists will remain unsatisfied.

For both Wallace and Dennett's accounts, there is a delicate issue concerning interpretation. Are these accounts supposed to capture and cohere with our ordinary beliefs and intuitions? Or, are we to understand these accounts as attempting to tell us about the property of responsibility, whatever its relationship to our commonsense concept? Here, I cannot do the full exegetical work required to defend one or the other interpretation. So, I will begin by assuming that whatever else the accounts are committed to, at the very least they are intended to capture and cohere with our ordinary beliefs and intuitions about responsibility. I take it that this assumption is in keeping with P. F. Strawson's original project of trying to account for "what we mean, i.e., of *all* we mean" by responsibility.[14] However, in accepting the assumption that these accounts are supposed to capture the contents of our ordinary beliefs about responsibility, this does not mean that the accounts cannot or are not also intended to be accounts of the truth conditions of responsibility. This only means that the accounts cannot be committed to providing truth conditions for responsibility irrespective of our folk beliefs. I will return to the significance of this assumption in part three.

B. WALLACE

Wallace's main innovation is something he calls the *normative interpretation* of responsibility. On the normative interpretation, and *pace* Strawson, judgments of responsibility do depend on facts about whether an agent is truly responsible. These background facts (upon which our responsibility judgments rely) are facts about the fairness of adopting the distinctive stance of holding someone morally responsible. That stance is understood in terms of a characteristic psychology and its associated practices. The complex metaphysics of agency defended by many incompatibilists turns out to be unnecessary, because the background facts are primarily facts about the fairness of a certain way of treating other

agents. And, these facts do not require an incompatibilist metaphysics of agency.

Wallace characterizes the normative interpretation of responsibility in the following way:

(*N*) *S* is morally responsible (for action *x*) if and only if it would be appropriate to hold *s* morally responsible (for action *x*).[15]

The first thing to note is that N gives an answer to the cognitivist criticism that plagued Strawson's theory. N is consistent with there being facts about whether we are truly responsible, and those facts can be important for our judgments of responsibility. This gives the language of responsibility a cognitivist construal, but the view retains the Strawsonian spirit of analyzing the concept of responsibility because the stance of holding responsible is understood in terms of a characteristic psychology and its associated practices.

Also worth noting is that N includes no specification of appropriateness. As it turns out, the details of Wallace's theory depend on understanding the appropriateness of N in terms of fairness. This yields a slightly different specification of the normative interpretation that we can call F:

(*F*): *S* is morally responsible (for action *x*) if and only if it would be *fair* to hold *s* morally responsible (for action *x*).

Though the move from N to F does not receive much discussion in *Responsibility and the Moral Sentiments*, F is a natural, though certainly contentious, refinement of N.

Wallace's acceptance of F turns on his view that the fairness (or not) of our responsibility-characteristic practices is, in some important way, prior to our being responsible. This assumption is not made clear by the schema itself, but it makes explicit what is innovative about Strawsonians. The difficulty, though, is that the innovation invites the charge that Strawsonians are failing to respect our commonsense understanding of responsibility.

Consider what is likely the standard view about the relationship between our being responsible and the appropriateness of our practices. Most incompatibilists and non-Strawsonian compatibilists believe that our being responsible is, roughly, a matter of an agent standing in a particular relation to an action. On this view, facts about responsibility are practice-independent facts about agency and action. If there is a relationship between responsibility and the appropriateness of our practices, facts about fairness depend on facts about being responsible. Call this the *agent-based account* of responsibility facts.

In contrast, Wallace maintains that our being responsible is *not* fixed by some facts antecedent to the appropriateness of our practices.[16] Rather,

an agent's being responsible depends on the fairness of treating that agent as responsible. Wallace's particular account might be described as a *normative practice-based account*. On this account, the "truth maker" for claims about responsibility is some normative feature of responsibility-characteristic practices (e.g., the fairness of the practices in general and/or in that specific instance). What makes it characteristically Strawsonian is that it is a member of the more general class of practice-based accounts, accounts where the truth maker is based on some feature of our practices.

As incompatibilists and other non-Strawsonians see it, Strawsonian accounts either misconstrue or fail to capture some core features of commonsense thinking about responsibility. The starting point for agent-based accounts is the idea that it is natural to think that responsibility facts are fixed by features of the agent and the agent's actions. This does not mean that responsibility-ascriptions cannot play other roles. Responsibility ascriptions may frequently play a dual role in our moral lives, both marking out facts about responsibility and indicating an assessment about the appropriateness of certain practices. But, what facts there are about responsibility are facts that supervene on agents and their actions, not on the practices directed at the agents.

As agent-based theorists see it, there are only two things we need to know to learn the facts about responsibility in any particular case: what kind of agent is involved, and the agent's connection to the considered action (or state of affairs). Wallace-style Strawsonians, however, maintain that we need to know a further thing: whether deployment of responsibility-characteristic practices is appropriate (or fair, etc.) in general and in the particular case. But, what evidence could they offer for thinking that we need to know these things as well? Everything we need to know seems to be settled by knowledge about the agent and his or her connection to the evaluated action or state of affairs. Indeed, the further normative facts of interest to normative practice-based theorists are, by their own admission, largely determined by facts about agents and their connection to actions or states of affairs. It seems gratuitous to insist that the normative property of being responsible is parasitic on a further, more basic normative property (e.g., the fairness of the practices), which is itself dependent on properties of agency and action on which the status of being responsible was initially thought to depend. Moreover, it seems possible to think that someone can be responsible, regardless of whether or not it is fair to hold them responsible in this or that particular case. Suppose we have a policy of never holding people responsible for their first moral infraction, even if we would normally be inclined to think of them as a fully responsible agent. Now suppose that we arbitrarily suspend this policy for a randomly selected person. It seems plausible to think of this as a case where it would be unfair to hold someone responsible all the while thinking that they are responsible. If so, this shows that

in our commonsense moral ontology, the property of responsibility is not dependent on some further and more basic normative property of our responsibility practices.[17]

Given the power of agent-based accounts to capture our commonsense thinking about responsibility, Strawsonians have to muster some compelling arguments to get incompatibilists and others to abandon the agent-based picture of our concept of responsibility. Wallace's main argument for a normative practice-based interpretation of our commonsense convictions is the fruitfulness of his account, which relies on this assumption.[18] But this kind of argument, especially given apparent counterexamples of the sort I mention above, is not likely to convince the majority of incompatibilists and non-Strawsonian compatibilists who find a practice-independent account of the folk concept more plausible. Thus, if Strawsonians want to sustain the claim that they can adequately capture folk thinking about responsibility, we need to look elsewhere for a defense of their compatibility with ordinary moral thinking.

C. DENNETT

Dennett's work is a promising place to look for a defense of practice-based accounts. Like Wallace, Dennett endorses a roughly Strawsonian interpretation of being and holding responsible. (Though unlike Wallace, he does not emphasize the importance of fairness in how we hold people responsible). More importantly, Dennett offers direct arguments against interpreting the concept of responsibility as having practice-independent purport.

Dennett rejects the agent-based picture of responsibility for three reasons. First, he thinks that taking the status of being responsible as prior commits you to a metaphysical interpretation where the invoked metaphysics cannot provide the requisite normative justification. Second, he argues that treating the status of being responsible as prior to holding responsible engenders intractable epistemological problems associated with responsibility. Finally, he thinks that there is no way to make the agent-based approach's presumed metaphysical story coherent.[19]

All of these reasons are inadequate for abandoning what Dennett admits is the common-sense assumption that we take *being* responsible as fundamental in our thinking about responsibility. For our purposes, the discovery that our folk metaphysics does not provide normative justification does not count as a reason for thinking that our concept has no such metaphysical commitments. It could well turn out that our ordinary concept of responsibility has metaphysical commitments that are not normatively justified. But, that would be a discovery about our concept, not something we should rule out as a matter of principle. The same is true of Dennett's third argument, about the coherence of the metaphysical story. Our folk metaphysical commitments might indeed be incoherent, but again,

that would be a discovery worth making, not a possibility we should close off at the start of inquiry.

As for Dennett's other charge, that a metaphysical grounding of responsibility would make assessments of guilt or innocence tricky things (inasmuch as facts of metaphysical independence would likely be epistemically inaccessible), it is difficult to see why this should count as a reason for thinking that our concept lacks metaphysical commitments.[20] Even if we can never demonstrably prove that someone has satisfied all the metaphysical conditions for true responsibility, Dennett's charge – at best – points to a need for practical ways of dealing with assessments of responsibility. But, this is hardly unique to the moral realm. We are always in need of practical solutions for problems about which we know we can never be certain of the best answer, even when we are confident that there is such a thing. Dennett himself makes this point in his 1988 Tanner lecture.[21]

In sum, neither Dennett nor Wallace offer arguments that would change the mind of an antecedently convinced incompatibilist or agent-based compatibilist. Thus, if Strawsonianism is to make good on its promise to end the pattern of dialectical stalemates, it will have to do so in a fundamentally different way it has attempted so far. In the rest of this paper, I attempt to show how this might be done.

3.

A. A CONCESSION AND TWO REACTIONS

The key to adequately rehabilitating the Strawson project and moving closer to a resolution of the deadlock between compatibilists and incompatibilists is for Strawsonians to concede something to incompatibilists. The concession is this: *the folk concept of responsibility may be incompatibilist*.

We can expect two different (and opposed!) reactions to this suggestion. The first reaction will go something like this: "You are groundlessly claiming that compatibilists should simply capitulate on the crux of the compatibility debate. Why in the world would any compatibilist agree to this?"

In response, it is important to consider what we (Strawsonian revisionists), are not giving up. We are not necessarily giving up on the idea that the *property* of responsibility is compatible with determinism, nor are we necessarily giving up on the idea that we can be responsible agents in a deterministic world. What we are giving up is the idea that an adequate theory of responsibility is one that fully captures folk beliefs about responsibility. In short, revisionist Strawsonians will admit that the best theory of responsibility might well be *revisionist* in the sense that it will depart (to some extent) from our commonsense understanding of responsibility, and ultimately, require some revision of commonsense. But, nothing in such a

concession requires that revisionist Strawsonians give up a commitment to the property of responsibility being compatible with the truth of determinism.

The kind of revisionism I propose is not altogether unheard of among Strawsonians. We find clear suggestions of it in Dennett's slogan of "the varieties of free will worth wanting" and in Wallace's conditional acceptance of "modest revisionism" about our retributivist folk beliefs, in light of the fairness demands imposed by F.[22] Moreover, there are some who have maintained that the only charitable way to interpret Strawson or compatibilists of any stripe is as revisionists.[23] If one already thought that revisionism was a central feature of compatibilism, the alternative reaction we can expect is something like this: "How is revisionist Strawsonianism any different than Strawsonianism? This is what Strawsonians have been trying to do all along. If what they have been doing isn't working, we should hardly expect that calling it revisionism will move us any closer to ending the dialectic of stalemates."

This reaction gives too much credit to extant Strawsonians. Despite sporadic awareness of it, the revisionist insight is almost never fully appreciated, even by those who admit it into their theories. Revisionism, when recognized at all, is usually admitted only cautiously and with some ambivalence.[24] For instance, suppose we read the texts of the aforementioned Strawsonians as arguments about the truth conditions for responsibility, where these accounts succeed or fail independently of the theory's conformity to the folk concept of responsibility. Such a reading does considerable violence to the structure of their texts. For example, there is a pervasive ambiguity in Dennett's account regarding whether the varieties of free will (or morally responsible agency) worth wanting are the one(s) we ordinarily do want. Sometimes, as in his analysis of control, he is concerned to give an account of "*our ordinary concept*."[25] At other times, when he considers intuitions in support of agent causation, for example, he does not argue that we do not have these intuitions or that we just need to understand their content properly. Rather, his aim is to give a naturalistically acceptable account of agency that does not rely on such intuitions, dismissing them as "a sort of cognitive illusion."[26] Similarly for Wallace; his admission of potential revisionism rests uneasily against a background of substantial argument directed at showing that we need not adopt metaphysical interpretations of our folk concept. If revising our folk concept of responsibility is acceptable in light of pursuing a normatively adequate account of responsibility, it is difficult to see why he should be concerned to undermine the intuitions that drive metaphysical accounts of responsible agency. It would seem better just to admit that we have a metaphysically demanding picture and then to argue that this picture should be abandoned in favor of the account he proposes.

A more thorough examination of the work of Strawsonians and other compatibilists would doubtlessly find more passages that are suggestive of

one or another form of revisionism.[27] To the extent that various Strawsonians intended to propose a revisionist project, what follows will already seem appealing. My goal, though, is to sketch how intentional, systematic, and rigorous pursuit of Strawsonian revisionism might constitute a genuine advance in our theorizing about responsibility.

I will begin by discussing the outlines of a general revisionist approach to responsibility. Then, I will argue for the advantages of a revisionism informed by Strawsonian insights.

B. OUTLINES OF A REVISIONIST PROJECT

Let us start by clarifying what the revisionist gives up to the incompatibilist. Suppose the revisionist concedes to the incompatibilist that our folk concept of responsibility really does suppose metaphysically demanding alternative possibilities, but that (for a variety of reasons), it is implausible to think that we have them. In principle, revisionists do not need to hinge their revisionism on alternative possibilities being a part of the folk concept. Revisionism could be adopted if our folk concept does not require alternative possibilities, but rather, some kind of agency that amounted to "unmoved mover-hood" or agent causation. As long as there is *some* incompatibilist condition required by the folk concept that is not likely to be met, there is room for a revisionist theory. Call the account of the (likely) unsatisfied incompatibilist condition the *folk conceptual error theory*.

Often there is an inclination to move from the conviction that the folk concept of responsibility is implausible to the conclusion that we are not responsible. It is important to note that such a move supposes a particular semantics of moral language. It supposes that reference to responsibility properties is largely or completely fixed by our concept of responsibility. But, we could hold a causal or some other externalist account of the reference of the relevant moral terms. If so, an error in our folk concept does not mean that we systematically fail to refer to some property of responsibility. It might only mean that we believe false things about responsibility. This insight, then, allows us to push past the stalemate between incompatibilists and compatibilists. We can concede that certain aspects of our thinking about responsibility are incompatibilist, without being committed to incompatibilism about the property of responsibility. It might well turn out that incompatibilism about the property of responsibility is true, too. But if an externalist semantics for responsibility is correct, then we will not learn this fact solely from reflection on our concept. The upshot is that we need not be held hostage to what we might call the *denotational content* of the concept.

Of course, there are surely many who would defend an internalist account of the reference of moral terms. For our purposes, though, we do not

need settle the issue one way or another. We can proceed with a fairly timid position: agnosticism about whether conceptual analysis tells us about the property of responsibility. Call this *semantic agnosticism*.[28]

Given acceptance of both semantic agnosticism and a folk conceptual error theory, how can the revisionist proceed? I propose to adopt two standards, perhaps hinted at in Strawson's own work, for the revision of the folk concept of responsibility. The first is a standard of *normative adequacy* and the second is a standard of *naturalistic plausibility*.

The standard of normative adequacy holds that however the revision goes, the result must include a concept that is justified and well integrated with our network of mutually supporting norms and practices. A revised concept of responsibility that made responsibility-characteristic practices immune to considerations of (for example) fairness, proportional praise or punishment, and differences of moral agency (from moral patients to fully moral agents) would hardly count as being well integrated. A revised concept of responsibility that played no justified normative role in our moral thinking, that systematically conflicted with other pieces of justified moral thinking, or that lacked normative force altogether would also fail to meet the standard of normative adequacy. Thus, the normative standard forces some degree of conservatism about the revision in order to preserve the normatively significant parts of our practices.

In order to satisfy the normative standard, a revisionist account will need to be justified independently of the non-revised concept of responsibility and concepts that depend on it. For instance, a revised concept would fail to count as justified if the attitudes and practices it is intended to preserve were justified solely in virtue of some normative notion that is itself conceptually dependent on the non-revised concept of responsibility. For example, if desert relies upon the non-revised concept of responsibility, then desert is an inappropriate basis for revising the concept of responsibility when we accept a folk conceptual error theory.

One might worry that the normative standard is problematic because it rules out justification involving concepts dependent on the current folk concept of responsibility. This might seem to deplete the stockpile of available normative concepts too much. However, it is not clear if the integrity of many, or any, normative concepts depends at all on the adequacy of the *concept* of responsibility. It is more sensible to think that the dependency would be on the property of responsibility. And, as we have seen, a folk conceptual error theory does not by itself entail an error theory about the property. Even if we admit that there are some normative notions whose justification or integrity depends on the folk concept of responsibility, there is no reason to think that their loss would significantly deplete the availability of normative concepts that might serve as a basis for revision. Fairness, virtuousness, rationality, and other important normative concepts seem to be underived from the folk concept of responsibility. Thus,

the justification of the bulk of our responsibility-characteristic practices and attitudes might come in terms of these notions.

This result is important when we consider the range of theories available to the revisionist. Recall schemas N and F:

(*N*) *S* is morally responsible (for action *x*) if and only if it would be appropriate to hold *s* morally responsible (for action *x*).

(*F*) *S* is morally responsible (for action *x*) if and only if it would be *fair* to hold *s* morally responsible (for action *x*).

Suppose we decided to construe these schemas as claims about the kinds of commitments our folk concept *should* have. In that case, the variation between N and F points to a variation between the kinds of normative claims that we are allowed to appeal to in a revisionist theory. In the case of N, as long as there is something that makes it appropriate to hold someone responsible (where we understand this as the distinctive stance of adopting the responsibility-characteristic attitudes and practices), we can justify those characteristic practices and attitudes. Appropriateness could, in principle, be decided in diverse ways, ranging from considerations of rationality, self-interest, and other values only contingently connected to morality. However, this is one place where it matters that the standard of normative adequacy restricts our revisionism in a particular way. If it turns out that the norms that justify the continuation of the bulk of responsibility-characteristic attitudes and practices are not, at some significant level, *moral* norms, it is difficult to see how what would be left could possibly count as moral practices and attitudes. If a revision of the concept of moral responsibility entails that the revised concept does not play the same sort of role (as a moral concept) in our network of norms, the revision will fail to meet the standard of normative adequacy. Hence, acceptance of the normative standard means our revisionist theory must be of a more specific sort than some allowed for by N.

The unsuitability of N suggests F as a candidate for understanding the constraints of a revisionist theory under the normative standard. On F, the revised concept of responsibility is restricted to justification solely in terms of fairness. However, the normative standard does not restrict theory as much as F proposes. To the extent that we accept that other moral notions survive acceptance of the folk conceptual error theory of responsibility, revisionisms based on these other justified moral concepts will meet the standard of adequacy. We might put things this way: a theory will count as satisfying the normative standard if it adheres to the following schema

(*M*) *S* is morally responsible (for action *x*) if and only if it would be *morally appropriate* to hold *s* morally responsible (for action *x*).[29]

Under this schema (and following Wallace's suggestion of what it is to hold someone responsible), the revisionist is committed to changing our folk concept of responsibility so that by "S is responsible" we understand *that there is some justified moral consideration or collection of considerations that entitles us to adopt towards* S *the stance characterized by those responsibility-characteristic beliefs, practices, and attitudes that are morally justified in a way not dependent on our current folk concept of responsibility.* Once the revision is firmly in place, when we say that "S is responsible" what we will have in mind is that our then-current responsibility-characteristic beliefs, practices, and attitudes concerning *S* are morally justifiable in light of whatever conditions the particular revisionist theory specifies.

The normative standard moves us closer to a plausible picture of revisionism, though a satisfactory revisionism will need to specify the particular conditions in light of which the bulk of responsibility-characteristic attitudes and practices are justified.

Let us turn to consider the standard of naturalistic plausibility. According to this standard, a revision must not require things that are implausible under some broad-minded conception of substantive naturalism. As I use it here, 'naturalism' need not be understood in an especially contentious way (e.g., as committed to strict reductionism).[30] Rather, we should think of it as helping to adjudicate a proposal's plausibility, based on what we know about science and the kinds of demands the considered theory makes on future science.[31]

We can see how the standard of naturalistic plausibility works in the following example. Suppose we learned that agent causation is scientifically implausible, if not impossible. In this case, commitment to the standard of naturalistic plausibility would prevent an agent causalist revision of our picture of responsible agency. Or, suppose we thought that the viability of a particular picture of agency depended on a very particular neurological structure, which we had no independent reason to believe in. Again, the naturalistic standard would treat this kind of theory as less plausible (ceteris paribus) than one that required no such structure.

What gives the naturalist standard some bite is that it effectively blocks a large class of theories from counting as viable revisionist accounts. Without some tool to reduce the total number of viable theories, admitting revisionism into our spectrum of theories might, by itself, seem to only double the number of accounts of the folk concept of responsibility. For any existing theory, we might suppose that it could be rendered in both revisionist and a non-revisionist ways. Thus, rather than ending the pattern of stalemates, revisionism might seem to make things worse. As is the case with the normative standard, adoption of the standard of naturalistic plausibility reduces the number of viable revisionist theories. In my judgment, though I will not attempt to adequately defend it here, it does this by ruling out virtually all revisionist theories that presuppose

libertarian agency. Given that we accept the need or possibility of conceptual revision, it would be prima facie undesirable to adopt a revision that makes significant demands on how the world must turn out (e.g., that indeterminism shows up in just the right place in the deliberative process and not in some other place, or that emergent causal powers appear at just the right level of ontological organization.).[32] Even though libertarians have made great strides in showing how their theories might be consistent with various forms of naturalism, it is an altogether different thing to convince us that the theories are naturalistically plausible. Moreover, it is a further challenge to show these are burdens we would want to impose on our revised concept. If we have a warrant for revising our folk concept in a number of different ways, why would we want to do it in a way that shoulders the burdens of libertarianism? Other than the already excluded motive of conserving our folk concept of responsibility, what motive could there be for putting our newly revised concept and its commitments at the mercy of speculative accounts of indeterminism?

C. REVISIONIST STRAWSONIANISM

Acceptance of a folk conceptual error theory and the standards of naturalistic plausibility and normative adequacy give considerable shape to a plausible revisionism. What we need, though, is some idea of how revisionists might go about filling in the indeterminate condition of moral appropriateness specified in M. In what follows, I sketch some of the ways in which reasonable revisionist theories can give some content to the schema provided in the previous section. My goal is not to argue for a particular account – this would be too much to attempt here. Rather, I hope to point out the ways in which development of a particularly Strawsonian revisionism will be well suited for giving an account of the moral appropriateness of some suitably large collection of our responsibility-characteristic beliefs, attitudes, and practices.

There are at least three reasons why specifically *Strawsonian* revisionism is promising. First, the revisionism sidesteps many of the complaints directed against Strawsonians. For example, Strawsonian revisionists need not deny those intuitions that suggest that our ordinary concept of responsibility is committed to alternate possibilities (or some other incompatibilist condition). All the favorite arguments of the incompatibilist can be accepted, as long as they are construed as arguments about our folk concept. This moves us closer to overcoming one stalemate with incompatibilists and it also gives new life to practice-based accounts of responsibility. For instance, even if practice-based accounts fail to fully capture the purport of the folk concept of responsibility, practice-based accounts may be appealing when recast as revisionist theories about the conditions for application of a revised concept of responsibility. That

means that any benefits of practice-based approaches to responsibility can be co-opted by revisionists. So, in the spirit of Strawsonianism, we might maintain that a practice-based account is the preferable way of avoiding the "panicky" metaphysical commitments created by our agent-based folk concept.

A second advantage of Strawsonian revisionism is that it can provide principled adjudication of debates in the theory of agency. Unlike standard compatibilists, revisionists need not worry whether their pictures of moral agency satisfy all of our pre-theoretical intuitions. We can expect that given the revisionist's focus on normative adequacy and naturalistic plausibility, revisionists will offer refinements of existing theories of moral agency that more closely track naturalist and normative standards than theories developed under concerns of intuitiveness. For instance, Fischer and Ravizza maintain that a theory of moral agency has to include a historical condition on the ownership of the agent's reasons-responsive mechanism.[33] Given the concern of preserving ordinary intuitions, this may be true. However, a rigorously revisionist approach might find little normative justification for retaining the historical condition. If so, this illustrates one way in which the very best parts of compatibilist theories might be re-deployed in the service of Strawsonian revisionism.

Finally, Strawsonian revisionists can benefit from traditional Strawsonianism's robust account of the moral psychology of holding agents responsible. For instance, if it turns out that Strawson was right that certain reactive attitudes and the practices they give rise to are genuinely inescapable, then the Strawsonian revisionist have an answer as to why at least those attitudes are justified or not in need of justification. Of course, there are likely to be complicated issues concerning the way belief revision affects other attitudes, and vice-versa.[34] But the point is that Strawsonian revisionism can help itself to all the available moral psychology. By incorporating these insights, revisionists can avoid accusations of psychological implausibility of the sort that have dogged some hard determinist theories.

Though this sketchy discussion of some advantages of revisionist Strawsonians is still some distance from a well-developed theory, we know enough to see some of the ways in which the Strawsonian revisionist can provide the foundations of a justification for the bulk of our responsibility-characteristic beliefs, attitudes, and practices. In particular, we can expect that the justification will be practice-based and tied to a revisionist-refined account of moral agency (whether hierarchical, reasons-responsive, or other). We can also expect the account to be informed by a robust account of the moral psychology of the attitudes and practices characteristic of responsibility.

Although the above-mentioned considerations are fairly abstract, it is enough to give us a method for developing concrete theories of responsibility. Here is the method:

1. Take your favored compatibilist theory of responsibility and invoke standard revisionist tropes (e.g., a folk conceptual error theory, the naturalist and normative standards, etc.) to justify the theory's partial departure from common sense.
2. Revise the theory's account of morally responsible agency so that it reflects our best picture of moral psychology.
3. Strip the account of morally responsible agency of any features that do not meet the naturalist and normative standards.
4. Show how the resultant specification of conditions for holding people responsible meets schema M.

There we have it – a method to build revisionist theories.

Of course, Strawsonian revisionism will not end every debate about free will and moral responsibility. Among revisionists there will be serious disputes about whether one package of revisions is more desirable than another. But, this just means that there is likely to be rich and fertile discussion between competing incarnations of revisionism. The chief advantage, though, is that these theories will be much better focused on what matters for responsibility and why.

D. REACTIONS TO REVISIONISM

Here I want to consider two different reactions, the first being the relationship of revisionism to normative ethics, the second being revisionism and incompatibilism.

Despite everything that has been said so far, one might reasonably wonder whether *any* kind of revisionism will be possible in the absence of a substantive moral theory. What a substantive theory of ethics gives us, among other things, is some account of the relations between various moral concepts and norms. Though the mentioned accounts of moral agency and psychology bring us closer to the specification of moral appropriateness that a revisionist needs to invoke, it might seem that at the end of the day we will still need a substantive theory of ethics to tell us what things are morally justified independent of our folk concept of responsibility. If so, then Strawsonian revisionism can only tell one part of the responsibility story.

I think this reaction is basically right. It seems implausible to think even a Strawsonian revision of moral responsibility can be done in a way that is *completely* independent of more general theories of morality. We should, however, be cautious about moving too quickly from the idea that there will necessarily be interaction between a theory of responsibility and a broader moral theory to the idea that a revisionist theory of responsibility should simply be the output of utilitarianism or virtue theory, for example.[35] Without additional arguments for the priority of

one kind of theorizing over another, we might even think the opposite: considerations grounded in a revisionist theory of moral responsibility will change the way we view broader theories of ethics.

For anyone agnostic about the truth of more general theories of ethics, or unsure about the appropriate direction of influence between theories of responsibility and broader theories of ethics, the best path will be to develop a theory of responsibility that is compatible with a wide array of plausible moral theories without being dependent on one in particular. We even have a model of how this might be done. Consider Wallace's own account of responsibility. It is guided by the idea of interpreting our practices and attitudes in terms of fairness. Now suppose a systematically revisionist recasting of his theory succeeds in justifying some sizeable subset of our practices in terms of a fairly thin notion of fairness. This proposal represent one example of how a revision might be justified on specifically moral grounds (i.e., fairness), without explicitly invoking a particular substantive theory of ethics.[36] The key seems to be starting with a fairly primitive and uncontroversial moral notion. Nonetheless, we should acknowledge that selection of nearly any moral notion, regardless of its "primitiveness," is likely to rule out one or another moral theory. The best we can hope for is an initial revisionist theory that makes relatively thin demands on a substantive theory of ethics.[37]

I will conclude by commenting on the relationship of revisionism to standard incompatibilist accounts. As we have seen, the clearest thing that revisionism offers incompatibilists is a willingness to concede that incompatibilist arguments do show that our folk concept of responsibility has incompatibilist commitments. In return, though, revisionists ask a high price: that we accept that our folk concept of responsibility should be revised so that it better conforms to the revisionist's interpretation of M. By the incompatibilist's lights, this price may well be too steep to pay. I think, however, that revisionism should be of concern for incompatibilists for at least two reasons.

First, libertarians ought to have an interest in revisionism for purely pragmatic reasons. Because libertarians generally understand their own proposals to be defeasible, they should have an active interest in what alternatives exist if libertarianism is falsified by future science or other means. The revisionism proposed might be treated as a more adequate "second best" theory of responsibility than standard compatibilist accounts.

There is also a philosophically deeper motive that should drive libertarians to care about revisionist proposals. Once revisionism is a viable theory, libertarians come under special pressure to say why we should not just pursue revisionism, regardless of one's favored view about determinism and responsibility. To the extent that revisionists are able to give a folk conceptually independent normative basis for the bulk of the beliefs, attitudes, and practices that are characteristic of responsibility, we are

forced to ask ourselves why we should want the libertarian network of concepts, practices, and attitudes over the revisionist's. Typically, libertarians do not feel a need to argue for their relative merits against compatibilist theories because compatibilists have been so quick to accept that an adequate theory of responsibility must fall within the constraints of the folk concept of responsibility.[38] As we saw in parts 1 and 2 of this paper, as long as this constraint is accepted by compatibilists (and inadequately met in the eyes of incompatibilists), incompatibilists will not feel compelled to answer challenges about why the beliefs, attitudes, and practices they account for are worth wanting. Libertarian freedom seems worthwhile at least because it is the only kind of theory that preserves our ordinary concept of responsibility. But, by calling into question the privilege of our folk concept, revisionists force libertarians to say why it matters so much that we preserve *all* the beliefs and attitudes characteristic of it.[39] Of course, libertarians and revisionists will disagree whether or not a folk conceptual error theory is likely to be true. The point, however, is that once we acknowledge that we can change some of our beliefs, practices, and attitudes in this domain, compatibility questions (including arguments about alternate possibilities) will be much less important than answers to why responsibility-characteristic practices, attitudes, and beliefs are important and worth keeping.[40]

Revisionism will also be of particular interest for more pessimistic forms of incompatibilism. Though the revisionism I present grows out of a compatibilist tradition, it is easy to see how, for example, hard deterministic theories might be re-written in revisionist terms. That is, we might take hard determinists to be arguing for a particularly stark version of revisionism. If such a reinterpretation is successful, that means that revisionism creates a bridge between the concerns of traditional compatibilists and the claims of pessimistic incompatibilists. Both kinds of revisionism would be joined in asking questions about what justifies the beliefs, practices, and attitudes that we have. If such convergence can be achieved, this would be no small accomplishment. Indeed, its mere possibility might be taken as a further reason to believe that revisionism provides a way to escape from some of our stalemates.[41]

Department of Philosophy
University of San Francisco

NOTES

[1] The term 'dialectical stalemate' is John Martin Fischer's. On a Fischer-related note, I should mention that my characterization of (in)compatibilism is meant to be neutral with regard to whether or not free will is required for responsibility. Thus, any references to free will will be treated as references to a kind of agency, condition, or power required for

morally responsible agency or moral responsibility. This device is not meant to reflect implicit acceptance of a substantive position on this issue (i.e., the relationship of responsibility and any or all notions of free will).

[2] P. F. Strawson (1962) "Freedom and Resentment," originally in *Proceedings of the British Academy*, xlviii, pp. 1–20 and reprinted (1982) in Gary Watson (ed.) *Free Will*, New York: Oxford, p. 59–80. For important developments and critiques, see Jonathan Bennett (1980) "Accountability" in Zak van Straaten (ed.) *Philosophical Subjects*, New York: Clarendon, Daniel Dennett (1984) *Elbow Room*, Cambridge: MIT, Gary Watson (1987) in Ferdinand Schoeman (ed.) *Responsibility, Character and the Emotions*, Ithica: Cornell, T. M. Scanlon (1988) "The Significance of Choice" in McMurrin (ed.) *The Tanner Lectures on Human Values* VII, Cambridge: Cambridge and his *What We Owe to Each Other*, Cambridge: Belknap Harvard, R. Jay Wallace (1996) *Responsibility and the Moral Sentiments*, Cambridge, Harvard, Michael Bratman (1997) "Responsibility and Planning" *The Journal of Ethics* 1, pp. 27–43, John Fischer and Mark Ravizza (1998) *Responsibility and Control*, Cambridge: Cambridge (1998) and Michael McKenna "The Limits of Evil and the Role of Moral Address" *The Journal of Ethics* 2, pp. 123–42.

[3] Further uses of the term 'responsibility' should be understood to stand for the more cumbersome term 'individual moral responsibility'.

[4] Op. cit., p. 260.

[5] Recall that consequentialist compatibilists were criticized for failing to take seriously the distinction between kinds of agents. Strawson's move seems to give the consequentialist a response: the distinction between responsibility practices and other kinds of practices *does* reflect a difference in moral agency precisely because the mark of moral agency is susceptibility to certain distinctive mechanisms of social and psychological influence. Only moral agents can be effectively influenced by considerations rooted in the characteristic attitudes and practices and this susceptibility is what marks them out as distinctive. Whether we should attribute this move to Strawson is perhaps disputable, but it is available to consequentialists of Strawsonian inspiration.

[6] Strawson later cites Carnap and Wittgenstein as the inspiration for this move – see Strawson's (1985) *Skepticism and Naturalism*, New York: Columbia, ch. 1–2.

[7] See Bennett, op. cit., Paul Russell (1992) "Strawson's Way of Naturalizing Responsibility" *Ethics* 102, pp. 287–302, and Watson, op. cit.

[8] Strawson admits that we can temporarily suspend the personal "reactive" framework that is partially constituitive of our inescapable social framework to "relieve the strains of commitment" that come from reactive engagement with others. However, he emphasizes that such suspensions are only temporary.

[9] Lawrence Stern (1974) "Freedom, Blame, and the Moral Community" *Journal of Philosophy*, pp. 72–84, Galen Strawson (1986) *Freedom and Belief*, Oxford: Oxford and Derk Pereboom (1995) "Determinism *al Dente*" *Nous* 29, pp. 21–45.

[10] Wolf, S. (1981) "The Importance of Free Will" *Mind* 90, pp. 386–405.

[11] Op. cit., p. 168.

[12] Berofsky, B. (2000) "Ultimate Responsibility in a Deterministic World" *Philosophy and Phenomenological Research* XL, p. 135.

[13] Both op. cit.

[14] Strawson, "Freedom and Resentment," p. 78. Though Bennett is inclined to read Strawson more along the revisionist lines I propose to make a systematic part of Strawsonianism in part 3, I think it is pretty clear that Strawson and many subsequent Strawsonians have been reluctant to see their project as conceptually revisionist. For example, consider that Strawson viewed libertarian metaphysics not as something given by conceptual analysis, but rather something postulated in an attempt to fill "a lacuna" in our

understanding of responsibility. The account Strawson proposed was therefore not intended as an attempt to overturn ordinary thinking, but rather, to show that our beliefs, practices, and attitudes were not committed to a very robust metaphysics, libertarian or otherwise. Of course, all of this is of a piece with Strawson's descriptive metaphysics.

[15] Op. cit., 91.

[16] Wallace opens his book by writing that, "if we wish to make sense of the idea that there are facts about what it is to be a responsible agent, it is best not to picture such facts as conceptually prior to and independent of our practice of holding people responsible." Later he reiterates the claim, emphasizing "on the normative approach, the facts about whether people are morally responsible are not yet available to be appealed to at this stage in the inquiry. Those facts are fixed by the answer to the question of when it is appropriate to hold people responsible, and so they cannot be invoked to decide that very question" Ibid, pp. 92–3.

[17] Randolph Clarke, who credits Scanlon as the source of his idea, suggested this case to me.

[18] In conversation, Wallace has confirmed that he does not take himself to have given a "direct" argument for the practice-based interpretation.

[19] Op. cit., p. 166.

[20] Note that if Dennett's argument did work, it would seem to work just as well against Wallace's account, for facts about fairness seem no more or less metaphysically spooky than facts about responsibility.

[21] See Dennett (1998) "The Moral First Aid Manual" in McMurrin (ed.) *The Tanner Lectures on Human Values* VII, Cambridge: Cambridge University Press. I take it there is more that can be said about this issue, though I do not think that it will change anyone's mind about these issues. For instance, Dennett could point out that our ordinary attributions seem to presume some confidence in assessments that ought to have at least some evidence. But the incompatibilist will just reply that we ordinarily assume people are metaphysically free unless we have *countervailing* evidence. What makes the issue of free will and moral responsibility an issue at all just is that it seems to threaten our unreflectively assumed belief that we have the kinds of metaphysically demanding powers we ordinarily assume. It is worth noting, though, that we should not simply assume that our current stock of philosophically interesting concepts is optimal, accurate, metaphysically innocuous, or worth keeping. More on this in section 3.

[22] Wallace, pp. 228–9.

[23] Bennett (1980) and Frank Jackson (1998) *From Metaphysics to Ethics*, New York: Oxford University Press.

[24] Bennett, op. cit., is an important exception. In that paper he (implausibly, I think) claims that Strawson's original theory was "excisionary." Despite important differences with Strawson he seems to endorse something like the revisionist strategy I have proposed here. However, his recent statement of a particular methodological approach in his (1995) *The Act Itself*, New York: Oxford University Press, and the way he links that approach to his earlier work make a straightforward interpretation of Bennett's project too complicated to pursue here.

[25] Op. cit., n.52, italics in original.

[26] Ibid, p. 77.

[27] For example, Fischer and Ravizza pursue a theory of responsibility that might be thought of as "revisionist" in the sense that it holds that serious reflection about our concept of control makes us realize that it need not be as metaphysically demanding as we initially suppose it to be. As I understand it, they think that our concept of responsibility-relevant control really is metaphysically innocuous, though they admit that there are intuitions that initially suggest otherwise. This relatively mild form of revisionism contrasts

with, for example, Wallace's and T. M. Scanlon's revisionism about retribution. They cautiously claim that their accounts may altogether depart from the commitments of folk thinking.

[28] Just how the proposed revision is understood will be something for revisionist metaphysics and semanticists to decide. However, one might worry that their answers have important consequences for the revisionist project. On the one hand, if our folk concept does fix reference, one might criticize revisionist theories as failing to be theories of *responsibility*. I'm inclined to think this is not a large worry. First, to the extent that the proposed revision is conservative, preserving the bulk of responsibility-characteristic beliefs, practices, and attitudes should be enough to earn the right to claim to be a revisionist theory *of responsibility*. Moreover, if we really do lack responsibility, it is hard to see how the fact that a proposed package of concept, attitudes, and practices does not pick out the exact same property (and note, in this case, a non-existent property) counts as a reason to think the theory is inadequate. In this case we might think of the theory as a charitable "paraphrasing" of our metaphysics of responsibility. On the other hand, if our concept does not fix reference and something else does, then one might think that any revision should be guided by whatever it is that fixes references. With respect to this latter criticism, the semantic agnostic revisionist need not disagree. However, I think that even if we could specify the truth conditions for responsibility in a folk-conceptually independent way, there might still be other reasons for taking up the revisionist's questions. For example, it could well turn out that the property we were tracking and calling 'responsibility' is not normatively binding in the way we ordinarily suppose (and of course, this possibility must be allowed for once we separate the specification of the property from our conception of it). In that case, we could still be interested in trying to ground the cluster of practices, attitudes, and beliefs characteristic of our old understanding of responsibility even if none of these things were justifiable in turns of the features of actually responsible things. See Manuel Vargas (forthcoming) "The Revisionists's Guide to Responsibility" *Philosophical Studies*.

[29] In keeping with the argument thus far, I am assuming that the revisionist will give priority to the idea that being responsible is to be understood in an intersubjectivist way.

[30] See Strawson's own discussion of responsibility in light of naturalism in Strawson, *Skepticism and Naturalism*, especially ch. 1–2.

[31] In adopting this standard I am assuming that realism about moral properties is compatible with naturalism. Even if one rejects this assumption there are still two options. First, one might accept the standard of naturalistic plausibility, but take a different view about how to talk about the issue of a revisionist picture of moral responsibility. Second, one might pursue a kind of revisionism compatible with only the standard of normative adequacy. Either way, these would count as revisionist theories, though not of the sort that I pursue.

[32] For a sophisticated account of the former, see Robert Kane (1996) *The Significance of Free Will*, New York: Oxford University Press. For sophisticated version of the latter, see Timothy O'Connor (2000) *Persons and Causes*, New York: Oxford University Press. Again, I do not mean to suggest that my remarks constitute adequate criticism for either theory. Rather, the point is merely to illustrate the greater burden of ontological commitment typically carried by libertarian theories.

[33] Op. cit.

[34] A complicated issue concerns the role of the cognitivist criticism in light of this possibility. In particular, one might worry that this argument suggests an abandonment of the cognitivist criticism. It need not, though. The revisionist will point out that even if our attitudes are inescapable, and even if those attitudes presuppose false or implausible beliefs,

those false beliefs are not the basis of justification in a revisionist theory. So even if we cannot, as a matter of everyday human psychology, fully replace folk beliefs with suitably cleaned up "revisionist beliefs," the revisionist will insist that our theorizing about responsibility reflect the beliefs we *ought* to have. In this way, a revisionist theory of responsibility might be a bit like some theories of physics which might be implausible candidates for replacing folk thinking, but true and theoretically necessary all the same.

[35] For a brief exploration of revisionism in light of virtue theory, see Michael Slote (1990) "Ethics Without Free Will" *Social Theory and Practice* 16, pp. 369–383. For a classic, but sometimes misunderstood statement of utilitarian revisionism, see J. J. C. Smart (1961) "Free Will, Praise, and Blame" *Mind* 70, pp. 157–163. On Smart's view, practices ought to be reorganized around procedures that promoted the good, which required a notion of praise and "dispraise," distinct from ordinary praise and blame. Smart's view is unusual in that he explicitly endorses a revisionist approach to moral blame.

[36] The revisionist Strawsonianism I propose departs from Wallace in at least three important ways. First, it is explicitly revisionist. It may also be revisionist to a greater degree than Wallace would accept. Second, it is open to grounding the normative integrity of our practices in terms other than just fairness. Third, it is strongly concerned to satisfy a standard of naturalism.

[37] I say "initial revisionism" to allow for two possibilities. The first is that as we learn more about what ethical systems are preferable, we will have reasons to advocate different kinds of revisions. This might motivate several rounds of "stages" of revisionism. Second, as the particular facts of our circumstances change what practices and attitudes are justifiable, we will have reason to call for more revisions in our folk concept of responsibility.

[38] One libertarian who has been a notable exception to this criticism is Kane, op. cit.

[39] Once this issue is opened up, revisionists will already be standing on the high ground because their theories are necessarily out to capture everything that is genuinely normatively binding and justifiable about our responsibility characteristic practices. By contrast, libertarian theories will start at a disadvantage because they will have been constructed to capture all of our ordinary intuitions and practices of responsibility. And, unless they can show the prima facie implausible, there is no reason to suspect that our current folk concept and practices track *only* what is plausible and justifiable.

[40] Part of the intractability of the incompatibility debate concerns an unarticulated difference in the range of ends that are considered primary for a theory of responsibility. For many, the appropriate end of a theory of responsibility is to provide a philosophical account consistent with our folk concept. For those engaged in "descriptive metaphysics" questions regarding the normative adequacy of these beliefs or categories are secondary, if they have any status at all. Opposed to the purely descriptive character of the metaphysical approach are theorists who also want to invoke concerns continuous with moral theorizing in general. These theorists include among the aims of a theory of responsibility a defense of the normative adequacy of the concept, attitudes, and/or practices constituitive of responsibility. Revisionism allows us to appreciate the truths of both projects.

[41] I owe thanks to many people for useful comments and valuable criticisms on this paper since it was first written, including Randolph Clarke, John Martin Fischer, Nadeem Hussain, Miriam McCormick, Michael McKenna, Derk Pereboom, John Perry, Tamar Schapiro, Ken Taylor, R. Jay Wallace, participants at the 2001 Inland Northwest Philosophy Conference, and Al Mele for his excellent referee remarks on this paper. Special thanks to Michael Bratman, Agnieszka Jaworska, and Michael McKenna for their considerable help with this paper and these ideas over the years.