Moral Sense and the Foundations of Responsibility

Paul Russell

The Oxford Handbook of Free Will: Second Edition (2 ed.)
Edited by Robert Kane

Print Publication Date: Jul 2011 Subject: Philosophy, Metaphysics
Online Publication Date: Sep 2012 DOI: 10.1093/oxfordhb/9780195399691.003.0010

Abstract and Keywords

This article discusses another important class of new compatibilist theories of agency and responsibility, frequently referred to as reactive attitude theories. Such theories have their roots in another seminal essay of modern free-will debates, P. F. Strawson's "Freedom and Resentment" (1962). This article disentangles three strands of Strawson's argument—rationalist, naturalist, and pragmatic. It also considers other recent reactive attitude views that have attempted to remedy flaws in Strawson's view, focusing particularly on the view of R. Jay Wallace. Wallace supplies an account of moral capacity, which is missing in Strawson's view, in terms of an account of what Wallace calls "reflective self-control." The article concludes with suggestions of how a reactive attitude approach to moral responsibility that builds on the work of Strawson, Wallace, and others might be successfully developed.

Keywords: compatibilist theories, reactive attitude theories, free will, P. F. Strawson, R. Jay Wallace, reflective self-control, moral responsibility

> It is a pity that talk of the moral sentiments has fallen out of favour. The phrase would be quite a good name for that network of human attitudes in acknowledging the character of which we find, I suggest, the only possibility of reconciling these disputants to each other.
>
> —P. F. Strawson, "Freedom and Resentment" (1962)

THROUGHOUT much of the first half of the twentieth century, the free-will debate was largely concerned with the question of what kind of freedom was required for moral responsibility and whether the kind of freedom required was compatible with the thesis of determinism. This issue was itself addressed primarily with reference to the question of how freedom is related to alternative possibilities and what the relevant analysis of "could have done otherwise" comes to. The discussion of these topics made little advance on the basic strategies and positions already developed and defended on either side of the compatibilist/incompatibilist divide in the preceding two centuries. When P. F.

Strawson's published his seminal article “Freedom and Resentment” in 1962 the dynamics of this debate were fundamentally altered. This is true both in respect of Strawson's general methodology, which demands a more empirically informed approach, and in terms of his core conceptual framework, which identifies a different set of considerations and issues at the heart of this debate. In particular, whereas the traditional or classical debate focused (p. 200) on the problem of (moral) freedom, Strawson directed his attention to the role of moral sentiments or “reactive attitudes” as the key to understanding and resolving the core problems lying at the heart of this debate. This essay is devoted to a critical assessment of Strawson's project and an analysis of the current debate concerning its prospects.

Strawson on Free Will and Reactive Attitudes

Strawson distinguishes two main camps in the free-will dispute, labeling them “optimists” and “pessimists” respectively. (Hereafter I will use these terms with capitals, to indicate Strawson's more technical sense of these terms.) Optimists are compatibilists who hold that our attitudes and practices associated with moral responsibility would in no way be discredited or dislodged by the truth of the thesis of determinism. The Pessimist, by contrast, is the libertarian who holds that moral responsibility requires the falsity of determinism and the possession of some form of “contra-causal freedom” (P. F. Strawson 1962, 73, 74, 92). A third position distinguished by Strawson is that of the “moral sceptic,” who holds that our “notions of moral guilt, of blame, of moral responsibility are inherently confused” whether determinism is true or false. Strawson's aim is to “reconcile” the Pessimist and Optimist positions (72). Specifically, he aims to show that although the “Pessimist” is correct in holding that the Optimist's account of moral responsibility leaves out “something vital” (73), what is needed to fill the gap in the Optimist's account is not any form of the “obscure and panicky metaphysics of libertarianism” (93). Optimists are mistaken, Strawson maintains, in supposing that we can understand and justify our commitment to the attitudes and practices of moral responsibility simply in terms of a “one-eyed utilitarianism” that is exclusively concerned with the social benefits of these practices (73, 92). Pessimists are mistaken in supposing that our commitment to these attidues and practices rests on the assumption that determinism is false.

Granted that the foundations of moral responsibility do not, on Strawson's account, rest with either libertarian metaphysics or consequentialist considerations regarding the social benefits of these attitudes and practices, where are we to discover the relevant foundations for moral responsibility? Strawson's strategy is to take what may be described as a “naturalistic turn.” Rather than asking directly, in the abstract, what *is* a responsible agent, Strawson suggests that we should consider in more detail, with more precision, what is involved in the attitudes that we take toward those who we regard as responsible agents. That is to say, what is involved in *holding* a person responsible? An approach of this kind depends less on a conceptual analysis of “freedom” and more on a descriptive psychology of human moral emotions. According to Strawson (1962, 75), our investigations in this area must begin with a basic fact about (p. 201) human beings: “the very

great importance we attach to the attitudes and intentions towards other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions." In this way, the correct starting point is to be found in "that complicated web of attitudes and feelings which form moral life as we know it" (91). When we proceed on this basis, we will place appropriate emphasis on the importance of (human) emotion in moral life and avoid the temptation—common to both Optimist and Pessimist strategies—to "overintellectualize" the free-will debate (91).

Two claims are fundamental to the Pessimist/skeptical view in the free-will debate. The first is that if the thesis of determinism is true, then we have reason to reject and repudiate the attitudes and practices associated with moral responsibility on the general ground that they are unjustified or incoherent. The second claim is that if we do indeed have reason to suspend or abandon the attitudes and practices associated with moral responsibility, in light of these skeptical reflections, then we are, psychologically speaking, *capable* of doing this. Strawson's central arguments in "Freedom and Resentment" are directly targeted against these two main prongs of the Pessimist/skeptical position. I will distinguish the two basic arguments in question as Strawson's "rationalist" and "naturalist" arguments. The first aims to show that the truth of determinism would not, in itself, systematically discredit our reactive attitudes and feelings, as associated with moral responsibility. The second aims to show that even if, contrary to what Strawson supposes, we are persuaded by the skeptical challenge, it is psychologically impossible for us to entirely abandon or wholly suspend our reactive attitudes on the basis of a "general theoretical conviction" of this kind (P. F. Strawson 1962, 81, 82, 87). In other words, as Strawson argues elsewhere, our *natural* commitment to the reactive attitudes insulates them against any form of global skeptical challenge (P. F. Strawson 1985, ch. 2).

Both the rationalist and naturalist components of Strawson's efforts to refute the Pessimist are presented in the framework of his analysis of the rationale of excuses. The Pessimist/skeptic maintains that if determinism is true, excusing considerations will (somehow) apply to all human action or hold universally. Specifically, according to the Pessimist/skeptical view, if determinism is true then we must systematically withdraw and suspend our reactive attitudes—collapsing our commitment to the entire edifice of moral responsibility. Under what circumstances, Strawson asks, do we "modify" or "mollify" our reactive attitudes or withhold them altogether? There are, he suggests, two different categories of excusing consideration (P. F. Strawson 1962, 77–79). The first, which I will refer to as excuses in the strict or narrow sense, do not imply that the agent concerned is an inappropriate target of reactive attitudes, or someone of whom we cannot demand some relevant degree of good will or due regard (77–78). In cases of this kind (e.g., accidents, ignorance), "the fact of injury [is] quite consistent with the agent's attitudes and intentions being just what they should be" (78). The features that concern us relate to the proper interpretation of the action or injury (e.g., that it was accidental, unintentional, lacked any ill-will). When we turn to the second category, what I will refer to as "exempting considerations," we are invited to withdraw or withhold entirely our **(p. 202)** reactive attitudes in respect of the agent. The exemption suggests that in some way the agent con-

cerned is not an appropriate target of reactive attitudes and not someone of whom we can make the usual demand of good will. Agents of this kind are judged inappropriate targets of our reactive attitudes, and our associated retributive practices, because they are either psychologically abnormal or morally underdeveloped (e.g., mentally ill, immature).

This analysis of the rationale of excuses allows us to see more clearly, Strawson claims, what has gone wrong with the traditional free-will debate. Granted that the issue of moral responsibility should be interpreted in terms of the conditions under which we view others as targets of reactive attitudes, would the truth of determinism "lead to the repudiation of all such attitudes"? (P. F. Strawson 1962, 80). Strawson answers, first, that the truth of determinism in no way serves (theoretically) to discredit our reactive attitudes in any systematic way. For this to be so, determinism would have to imply that one or other of the two basic forms of excusing considerations hold universally. There is, according to Strawson, no reason to believe that this is the case. Clearly determinism does not imply that every injurious action is done accidentally or unintentionally. Determinism does not imply that no one's conduct ever manifests ill will or fails to show proper regard for others. Nor does determinism imply that all agents are somehow "abnormal" or immature (81). It follows from these observations that the truth of determinism in no way discredits or theoretically undermines our commitment to reactive attitudes of the kind involved in our ascriptions of responsibility. Contrary to what the Pessimist/skeptic maintains, therefore, the truth of determinism does not erode the necessary metaphysical foundations of our attitudes and practices associated with moral responsibility.

With this rationalistic argument in place, Strawson proceeds to support his critique of the Pessimist/skeptical position with his naturalist argument. Even if we had some theoretical reason to entirely abandon or suspend our reactive attitudes (e.g., as per the skeptical challenge), it would be psychologically impossible for us to do this. To do this would involve "adopting a thoroughgoing objectivity of attitude to others," which is something Strawson claims we are incapable of (P. F. Strawson 1962, 81–30; 1985, 39).

> To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something … to be managed or handled or cured or trained…. But it cannot include the range of feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships … (P. F. Strawson 1962, 79).

Strawson allows that there are two circumstances in which the objective attitude is available to us. First, in circumstances where exempting conditions apply (e.g., mental illness), the objective attitude is, in fact, *required* of us, insofar as we are "civilized" (P. F. Strawson 1962, 81–82). There are also circumstances when the objective attitude may be adopted towards a "normal and mature" person simply because we want to use it as a refuge from "the strain of involvement" (82). However, (p. 203) Strawson is careful to emphasize the limits of any policy of this kind. Although it is necessary to adopt the objective attitude towards those individuals who are "abnormal or immature," and although it is al-

so possible to extend this attitude to some normal people on some occasions, a "sustained objectivity of inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth [sc., determinism] were a theoretical ground for it" (81). In other words, according to Strawson, the Pessimist/skeptic cannot *live* his skepticism—from a practical point of view skepticism of this kind is irrelevant (see, especially, P. F. Strawson 1985, 38–39). In face of the skeptical challenge, Strawson's naturalistic riposte is to claim that "it is *useless* to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do" (P. F. Strawson 1962, 87; emphasis in original). In this way, the skeptical challenge, based on worries about determinism, is not only groundless, it is also useless and irrelevant, because it has no potential practical or psychological traction in human nature and human life.

Beyond his rationalist and naturalist arguments, Strawson adds a third argument, which we may call his "pragmatic argument." Even if, contrary to the naturalistic observations that have been advanced, we were to suppose that we might be given a "god-like choice" concerning whether we should abandon or retain our (natural) commitment to the reactive attitudes, this choice, Strawson argues, must be decided in terms of the "gains and losses to human life, its enrichment or impoverishment" (P. F. Strawson 1962, 83). Clearly, on Strawson's account, any choice to abandon or altogether suspend our commitment to reactive attitudes would involve trying to live our lives from entirely within the "objective" stance—something that would imply total "human isolation" and a bleak, dehumanized existence (81, 83, 89, 93). As Strawson presents it, any (notional) choice that we may be in a position to make concerning whether to continue to participate in a social community of human relationships, constituted and held together by our reactive attitudes, cannot and should not be decided with reference to a "theoretical" issue such as determinism. On the contrary, because our commitment to the reactive attitudes is, on this account, essential to our very humanity, no sane or sensible person would linger long over this question, even if it were to be presented to us.

Although Strawson's principal arguments are directed against the Pessimist's skeptical view, he draws important conclusions from these arguments that make clear how his own compatibilist position diverges from that of the Optimist or classical compatibilism. The Optimist generally attempts to show that the truth of determinism does not prevent rewards and punishments from "regulating behavior in socially desirable ways" (P. F. Strawson 1962, 89):

> The picture painted by the optimist is painted in a style appropriate to a situation envisaged as wholly dominated by objectivity of attitude. The only operative notions involved in this picture are such as those of policy, treatment, control. But a thoroughgoing objectivity of attitude, excluding as it does the moral reactive attitudes, excludes at the same time essential elements in the concepts of *moral* condemnation and *moral* responsibility (89; emphasis in original).

(p. 204) The Pessimist is right, Strawson argues, to "recoil" at this picture of things but makes the mistake of assuming that "the gap in the optimist's account … can be filled only if some general metaphysical proposition [sc., indeterminism or contracausal freedom] is repeatedly verified, verified in all cases where it is appropriate to attribute moral responsibility" (92).

According to Strawson, in the final analysis, both the Optimist and the Pessimist are guilty of a shared misunderstanding:

> Both seek, in different ways, to over-intellectualize the facts. Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes is itself something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external "rational justification." Pessimists and optimists alike show themselves, in different ways, unable to accept this (P. F. Strawson 1962, 91–92).

It is, evidently, no part of Strawson's view to suggest that the reactive attitudes are altogether incapable of (rational) justification and criticism. On the contrary, Strawson's remarks explicitly make the point that "inside the general structure or web of human attitudes and feelings" there is a place and role for justification and criticism—this being an obvious corollary of his analysis and observations relating to the rationale of excusing and exempting conditions. The important point remains, however, that the basis of our general commitment or liability to the reactive attitudes is not itself something in need or capable of any form of theoretical or practical justification. These are emotional dispositions rooted in human nature at a deeper level than that provided by any (unconvincing and unnecessary) philosophical justifications.

Having identified the relevant lacunae in the Optimist's position, and the faulty alternative analysis provided by the Pessimist, Strawson (1962, 93) goes on to conclude that "if we sufficiently, that is radically, modify the view of the optimist, his view is the right one." Reconciliation can in this way be achieved when we take note of the fact that our retributive practices "do not merely exploit our natures, they express them" (93). When this core insight is fully appreciated, and the gap in the Optimist's position has been filled, there is no need to fall back into "the obscure and panicky metaphysics of libertarianism" (93).

Assessing Strawson's Arguments

Having described the core arguments that feature in Strawson's strategy, we may now assess them for their strengths and weaknesses (see also Haji 2002a; Kane 2005a, ch. 10). Each of the three core arguments we have described—rationalist, naturalist, (p. 205) and pragmatic—encounter serious difficulties, if they are not fatally flawed. Let us consider, first, Strawson's rationalist argument. The key objective, for the success of this argument, is to show that, even if determinism is true, none of the standard excusing and ex-

empting conditions can be generalized or said to hold universally (i.e., in virtue of the truth of this metaphysical thesis). Specifically, a crucial aspect of this argument involves showing that we have no reason to suppose, contrary to the Pessimist/skeptic, that exempting conditions apply to everyone if determinism is true. Critics, as well as some followers of Strawson, have found his argument unconvincing (see, e.g., Nagel 1986, 124–26; Watson 1987a, 262–63; Russell 1992). According to Strawson (1962, 81),

> the participant attitude, and the personal reactive attitudes in general, tend to give place, and it is judged by the civilized should give place, to objective attitudes, just insofar as the agent is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality—or simply by being a child. But it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition.

The weakness in this argument is that it plainly equivocates between “abnormal” and “incapacitated.” Contrary to what Strawson's language suggests, it is incapacity, and not abnormality, that serves as the relevant basis for exemptions. This leaves his anti-skeptical position open to a direct rejoinder from the Pessimist/skeptical camp.

The Pessimist/skeptic should not be understood as claiming that if determinism is true we are all abnormal. Rather, the Pessimist/skeptic claims only that if the thesis of determinism is true, then we are all incapacitated and, consequently, inappropriate targets of reactive attitudes. There is nothing self-contradictory about a thesis that suggests that incapacity is a universal condition. The relevant capacity, according to those Pessimists who accept libertarian metaphysics, is “free will” or “contra-causal freedom” of some kind. P. F. Strawson (1962, 93), as we have noted, maintains that this view would commit us to “obscure and panicky metaphysics” and imposes upon us a condition of responsibility “which cannot be coherently described” (P. F. Strawson 1980, 265). Even if Strawson is right about this, his response does not show that the thesis of determinism poses no threat to our moral capacities and, hence, to our reactive attitudes as a whole. At most, all Strawson succeeds in doing is casting doubt on one interpretation of what the relevant capacities are supposed to be. What we require, however, in order to discredit the skeptical threat, is an account of what is involved or required of our moral capacities, such that we can say who is or is not exempted of responsibility (i.e., who is an appropriate target of reactive attitudes). Without some more plausible and detailed alternative characterization of the nature of moral capacity, we are in no position to give assurance that the truth of determinism is irrelevant to this issue. Although something of an appropriate nature can, perhaps, be said on behalf of the rationalist argument, we cannot find it in his own remarks on this subject (Russell 1992, 153–55). We may conclude, therefore, that Strawson's reply to the Pessimist/skeptic is, at best, incomplete.

(p. 206) What, then, can we say about Strawson's naturalistic argument? The difficulties Strawson faces here are, if anything, even more severe and fundamental. The key to Strawson's naturalistic response to the Pessimist/skeptical challenge is to claim that our commitment to the whole framework or web of the reactive attitudes does not require any

kind of general rational justification and that no general “theoretical conviction” is capable of entirely dislodging this commitment. Nothing of this kind can lead us to repudiate all our reactive attitudes. Considered as a way of refuting or discrediting Pessimism and skepticism, Strawson's reply relies on two different forms or modes of naturalism, which Strawson fails to distinguish. Strawson's remarks suggest that he reads the Pessimist/skeptic as demanding some general rational justification for our liability or proneness to reactive attitudes. It may well be correct to claim, as Strawson does, that our liability to these emotions, as a type, is a natural fact about us that neither requires nor is capable of any rational (philosophical) justification. So considered, Strawson advances what I will call a “type-naturalist response” to the skeptical challenge. A response of this kind cannot, however, deal adequately with the Pessimist/skeptical threat properly understood.

The Pessimist/skeptic should be understood as claiming only that, given the truth of determinism, we are never justified in entertaining (any) tokens of reactive attitudes. In other words, however prone or liable to reactive attitudes we may be, in these circumstances, praising and blaming are never appropriate or legitimate. This form of skepticism—as it concerns tokens of reactive attitudes—is perfectly consistent with accepting Strawson's type-naturalism. Although we may be naturally prone or liable to these (moral) emotions, we are nevertheless capable of ceasing to feel or entertain these emotions if and when we judge, in the relevant circumstances, that these emotions are unjustified. The only naturalist reply to this (distinct) form of Pessimism/skepticism is to insist that no reasoning of any sort could ever lead us to cease entertaining or feeling emotions of this kind. Whatever considerations are brought to our attention regarding our human predicament—whatever reason may suggest to us—we will nevertheless continue to experience and feel emotions of this kind (i.e., tokens of this type of emotion). This form of token-naturalism is, psychologically speaking, less plausible than its type-naturalist counterpart, because it is not evident that our (token) emotional response cannot be controlled by reason and reflection when we judge that these emotions are inappropriate and uncalled for. From another point of view, token-naturalist claims, even if they are accepted, would do nothing to refute or discredit the core Pessimist/skeptical objection and worry—which is that if determinism is true our reactive attitudes are never justified or legitimate. Even if it were true that we are, in some way, constitutionally incapable of ceasing to entertain these emotions, this would not serve to address the relevant justificatory issue that is the focus of the Pessimist/skeptical challenge. Indeed, if our commitment to the fabric of responsibility rests, in the final analysis, on a *token*-naturalist psychology, this is more disturbing than simple skepticism—because it implies that attitudes and practices that we recognize as reflectively unjust and inappropriate cannot be brought under (p. 207) the control of reason. A naturalism of this kind is as unattractive as it is implausible (Russell 1992, but see Nichols 2007a).

Strawson's third core argument, his pragmatic argument, is likewise misguided and unconvincing. His type-naturalist claims about our natural liability or proneness to reactive attitudes does lend support to his claim that there is no question of us making some “godlike choice” about whether to retain or dispense with our general disposition to the participant stance (see, e.g., P. F. Strawson 1985, 31–38). Nevertheless, if we were given this

godlike choice (i.e., relating to our commitment to this type of emotion), it does not follow we would be entitled to decide whether to entertain tokens of reactive attitudes on the basis of considerations relating to "the gains and losses of human life, its enrichment or impoverishment" (P. F. Strawson 1962, 83). On the contrary, should we be in a position to choose to retain this commitment—contrary to the type-naturalist hypothesis—we would still be constrained by the "internal" rationale of this commitment to suspend any and all tokens of reactive attitude where and when relevant excusing and exempting conditions apply. It follows from this that, if the Pessimist/skeptical challenge is well-founded, we cannot aim to justify *tokens* of reactive attitudes on the grounds that in their complete absence our lives would be somehow "impoverished" or "less human." If this were the case it would certainly be a bleak situation, but we cannot insulate ourselves from this (theoretical) possibility by simply setting aside the relevance of exempting considerations as they apply to the framework and coherence of the reactive attitudes (see, however, Nichols 2007a, who finds more mileage in the pragmatic dimension of Strawson's strategy).

Exemptions, Moral Capacity, and Reflective Self-Control

The assessment of Strawson's "reconciling project" provided above makes clear that his approach encounters serious and substantial difficulties. It would be wrong, however, to conclude that no further headway can be made by following the tracks Strawson has laid down. On the contrary, a sympathetic reconstruction of Strawson's project, avoiding some of the weaknesses, and filling-in some missing elements, may still provide a plausible alternative to libertarian metaphysics, utilitarian-oriented compatibilism, and moral skepticism—each of which have their own difficulties and flaws. The relevant starting point for such a project rests with a more robust and plausible account of moral capacity. The absence of a detailed account of moral capacity, as we have seen, is a major weakness in Strawson's own contribution and, in particular, leaves his rationalist argument open to objection (Russell 1992; on similar difficulties relating to Hume's theory see Russell 1995, ch. 6). If a more adequate theory of moral capacity is available, then Strawson's approach can (p. 208) be provided with an account of exemptions that will serve his compatibilist objectives. An important and influential attempt to supply Strawson's with these elements has been provided by R. J. Wallace.

In *Responsibility and the Moral Sentiments*, Wallace defends a compatibilist position that combines two strands of philosophical thought, a Strawsonian account of holding people responsible and a Kantian theory of moral agency. Methodologically speaking, Wallace (1994, 5–6, 15) presents himself as offering a "normative interpretation" of the free-will debate. In Wallace's account, it is crucial that we begin our investigations, as Strawson does, with a philosophically adequate description of what is involved in holding people responsible, because the conditions of responsibility must themselves be interpreted in terms of when it is *fair* for us to adopt the stance of holding an agent responsible (15–16).

In other words, Wallace is skeptical of any effort to describe conditions of responsibility in the abstract without reference to what is involved in holding an agent responsible. His account of holding responsible has, in this sense, priority over his account of being responsible.

Wallace's discussion makes substantial contributions that fall on either side of the Strawsonian and Kantian elements mentioned above. On the side of his Strawsonian account of holding agents responsible, Wallace provides a different taxonomy of the reactive attitudes in relation to the moral emotions from that suggested by Strawson. Wallace uses this taxonomy of the reactive attitudes and moral emotions to carve out a distinct and different set of commitments on the issues of "objectivity," "naturalism," and "pessimism"—all issues where he diverges significantly from Strawson. On the other side of the divide, his Kantian theory of agency, Wallace presents an outline of a theory of "reflective self-control" that provides a principled, normative basis for exempting conditions, consistent with his basic compatibilist ambitions.

Let us begin with the key elements of Wallace's account of holding people responsible. To hold a person morally responsible is "to hold the person to moral expectations that one accepts" (Wallace 1994, 51). In this view, moral expectations are supported by moral reasons or justifications, and expectations of this nature constitute obligations (36, 63–64). Moreover, there is an essential linkage between holding someone to a (moral) expectation and being susceptible to (moral) reactive attitudes, such as resentment, indignation, and guilt. Susceptibility to these emotions is, Wallace maintains, "what constitutes holding someone to an expectation" (21). This mutual dependence of emotion and expectation is what distinguishes the reactive attitudes (21).[1] One notable advantage of this general account of what is involved in holding a person responsible, presented in terms of the essential relationship between expectations and reactive attitudes, is that it enables us to provide a theory of reactive attitudes that has some cognitive content, as opposed to a cruder, emotivist understanding of reactive attitudes or moral sentiments understood merely as raw feelings (74–78; and see Russell 1995, ch. 6, as this issue relates to Hume's system.)

Wallace's analysis, although it clearly provides the Strawsonian system with some precision and detail, comes at some cost. One of the more obvious difficulties **(p. 209)** is that if we accept this account of holding agents responsible, it follows that we are committed to a "narrow" view of responsibility that focuses exclusively on negative emotions (e.g., indignation, resentment, and guilt, as aroused in circumstances when expectations/obligation are judged to have been violated). Wallace (1994, 63–64, 71) attempts to explain away this worry about his "asymmetrical" treatment of responsibility in terms of what he claims is the absence of any particular "positive emotions" in relation to morally worthy actions. Suffice it to say that this is an oddity of Wallace's account that is not present in Strawson's own contribution. Another, and perhaps more fundamental, difficulty that Wallace considers is the objection that we may—and often do—hold people responsible without engaging any particular emotion toward the person concerned (76, and also 23, 62). Wallace's reply to this objection is that although we must understand the stance of hold-

ing people responsible with reference to the relevant (moral) reactive attitudes, this does not require that “we actually feel the relevant emotion in all the cases in which it would be appropriate to do so” (76). What his theory commits us to is a disjunctive requirement that in holding a person responsible we must either be susceptible to the reactive emotions or believe that it would be appropriate for one to feel the reactive emotions, when the relevant expectations are violated (23, 62, 76). It follows from this that although feeling or engaged emotion is not required for holding a person responsible, some relevant connection with these emotions and feelings is still required (i.e., via the belief that they are appropriate).

Although Wallace is anxious to clear his theory of any emotivist or noncognitivist features, the essential connection between responsibility and the reactive attitudes remains fundamental to his account. Wallace (1994, 52) explains the importance of this connection —the foundations of moral responsibility in our emotions—in terms of the issue of “depth” (cp. Wolf 1990, 41). Without any reference to moral emotions and feelings of the kind Wallace has described, the force of moral judgments of blame and responsibility would be lost. True moral blame, he suggests, is a form of deep assessment that reflects or manifests an attitude toward the agent who has acted wrongly (Wallace 1994, 78). Any account that severs judgments of responsibility from the set of attitudes associated with them (i.e., the “distinctive syndrome” associated with moral assessment; see Wallace 1994, 24) would render blame “superficial” (78). For this reason, Wallace insists that judgments of responsibility must reach beyond a mere description of what the agent has done (e.g., violating our moral expectations) and account for the condition of the judge who assigns blame (81–83). It is the judge's stance that captures the attitudinal dimension that gives blame and our judgments concerned with moral responsibility their distinct force and depth. It is this feature of Wallace's position that explains why, on his account, our understanding of responsibility must begin with an adequate analysis of holding people responsible. However much Wallace's position diverges from Strawson's views in other respects, he remains faithful to this key feature.

Although the adjustments and modifications that Wallace makes to the Strawsonian side of his position are significant, his most important contribution rests with his Kantian account of moral agency and moral capacity. As we have **(p. 210)** noted, when we considered Strawson's rationalist argument, it is here that Strawson's position is at its weakest and most vulnerable. Wallace defends a theory of “reflective self-control” that is, as he presents it, a form of “practical freedom” of a recognizably Kantian kind (Wallace 1994, 12–15). To explain the nature and character of his conception of moral agency, Wallace distinguishes “two competing pictures of what it is to be a morally responsible agent” (86). The picture Wallace rejects is one that interprets “the apparent truism that moral responsibility involves a kind of control over one's action” in terms of possessing a causal power over a range of alternatives. In this picture, moral agency requires genuine alternatives—something that “invites an incompatibilist understanding of responsibility, as requiring strong freedom of will” (86). Another view of control over actions, however, is concerned with the possession of “normative competence.” Normative competence should be understood in terms of (1) the power to grasp and apply moral reasons, and (2)

the power to control or regulate behavior in light of such reasons (86, 157). Agents who have these powers are capable of "reflective self-control." Although determinism may deprive us of genuine alternatives, it does not necessarily deprive us of the relevant powers of normative competence that Wallace has described. (Other influential compatibilist accounts of rational self-control are found in Dennett 1984; Wolf 1992; Fischer and Ravizza 1998.)

On Wallace's (1994, 15) normative interpretation, the "conditions of responsibility are to be construed as conditions that make it fair to adopt the stance of holding people responsible." In light of this, the relevant question to ask is: Would the truth of determinism make it unfair to hold someone responsible, where this is understood in terms of directing reactive attitudes at someone who has violated the relevant moral expectations? Clearly, where ordinary excuses in the narrow sense apply we must withdraw or inhibit our reactive attitudes, because the point or force of excuses is to establish that "the agent did not really violate the moral obligations we accept after all" (133, 147). In other words, where valid excuses hold, the agent has done nothing wrong and there is, in fact, no fault to be found in the quality of the agent's will (135). We may account for considerations of this kind, Wallace argues, without reference to alternative possibilities or the need for "strong freedom of will." What, then, about exempting conditions? In Wallace's normative competence picture, it is fair to hold an agent responsible so long as she possesses the relevant powers of "reflective self-control." In the case of children or the insane it would indeed be unfair to hold them responsible, given that they lack these capacities for reflective self-control. Again, however, the relevant distinctions can be drawn here, Wallace argues, without relying on the metaphysics of indeterminism and (genuine) alternative possibilities (181). On this basis, Wallace concludes, a compatibilist view can be constructed and defended from within the constraints of the "normative interpretation of the debate about responsibility."

Although Wallace's account of Kantian agency lends considerable support to a broadly Strawsonian strategy, it remains vulnerable to a serious objection—one that Wallace anticipates but does not convincingly defuse. In Wallace's account, it is fair to hold a person responsible for doing wrong even though they may have been (p. 211) unable to exercise their powers of reflective self-control differently in the actual circumstances. All that is relevant to the question of the agent's responsibility, Wallace maintains, is that the agent possesses the relevant general powers (i.e., qua disposition) and in fact exercised those powers in such a way that the relevant expectations were violated (Wallace 1994, 161–62). The difficulty remains, however, that the mere possession of such powers does not give the agent control over *the way in which they are actually exercised* (on this see Kane 2002e; see also Russell 2002, 244–5 concerning related difficulties for Dennett 1984). Although Wallace worries over this problem (182–86, 196–214, 223), his position, in the end, reduces to his insistence that this further condition (i.e., that the agent can control how his powers are actually exercised) would simply "give the game away" to the incompatibilist (223). What the Pessimist/skeptic needs here, and will not find in Wallace's discussion, is a convincing account of why it is fair to hold a person responsible for conduct that flows from powers that are exercised in ways over which they have no control. With-

out a more substantial reply to this objection, the Strawsonian strategy that Wallace pursues will not persuade its critics.

Holding and Being Responsible

Wallace, as we have noted, makes clear that his effort to reconstruct the moral sentiments approach to responsibility, along the lines advanced earlier by Strawson, falls into two component parts: Strawsonian and Kantian. This division of labor looks essential to the viability of the entire project, because the theory of "holding responsible," on one side, requires a theory of "responsible agency" on the other. This divide is, however, problematic from several points of view. It may be argued, for example, that insofar as Wallace's Kantian theory of agency is judged a success, it is no longer evident that we need a "normative interpretation of responsibility" that supposes that conditions of responsibility are to be construed as "conditions that make it fair to adopt the stance of *holding* people responsible" (Wallace 1994, 15; emphasis added). That is to say, if we can provide a full and complete account of being a responsible agent in terms of agents possessing powers of reflective self-control and being subject to relevant moral norms, why must we include any reference to the role of reactive attitudes or moral feelings in this context? Such elements may be judged as not only unnecessary, but also misplaced and misleading. Criticism along these general lines has been developed by Angela Smith in her recent article "On Being Responsible and Holding Responsible."

Smith (2007, 466, 472, 483) argues that Wallace's normative interpretation, and by implication all similar Strawsonian strategies, confuse two distinct sets of issues and conditions. Specifically, there is a distinction to be drawn between: (1) the conditions under which it is fair and appropriate to blame people, and (2) the conditions under which it is appropriate to judge them to be responsible and blameworthy (p. 212) (472). Smith's account of this matter turns on a related distinction between the agent being "at fault" or "culpable" and it being fair to blame the agent (see, e.g., 466n5). Culpability or blameworthiness implies the agent is at fault and subject to (valid) criticism. It does not follow, however, from the fact that a person is at fault or culpable that "active blaming" is appropriate (473). Active blaming, as Smith understands it, "in some way goes beyond beliefs about a person's responsibility and culpability" (470). Smith grants that her terminology in this respect is potentially misleading if it is taken to imply that the "'active blamer' must actually *do* something to express her blame towards the person she blames" (477; emphasis in original). This is not necessary, because active blame may involve simply feeling resentment, indignation, or anger toward the agent, without expressing these emotions in any way. Nonetheless, although blaming presupposes culpability, culpability or fault does not, by itself, entail that blaming is appropriate (473n10; see the related discussion in Kutz 2000, ch. 2). This gap between conditions of culpability and appropriate blaming, Smith argues, shows that conditions of being responsible cannot be reduced to conditions of appropriate active blaming.

How, then, do we assess when it is appropriate to actively blame an agent for some fault or wrongdoing? The relevant variables here, according to Smith, include considerations such as (i) our own standing as possible or potential moral judges, (ii) the significance of the fault to which we are responding, and (iii) the nature of the agent's own response to the fault or conduct in question (Smith 2007, 478). In respect of all these issues, Smith claims, issues of culpability and appropriate blame come apart and may diverge. For example, I may regard myself as not standing in a relevantly close or intimate relationship with an agent to be in a position to actively blame him for a fault (e.g., treating his spouse in an inconsiderate manner), even though I may well judge the agent is at fault and culpable. Likewise, I may regard the fault or culpable conduct as too insignificant or unimportant to merit resentment or indignation without compromising the initial judgment that the agent is responsible for some wrongdoing. Finally, in some cases the agent's own response to her faults (e.g., her obvious remorse and guilt) may encourage the view that any active blame is uncalled for and inappropriate. Again, this conclusion may be reached without compromising our independent and distinct judgment regarding her responsibility or culpability for her actions. With respect to variables and considerations of these kinds, because they concern the conditions of when it is appropriate to actively blame a person, there "may be no single, definitive answer to this question, because the 'us' in question [i.e., qua moral judges] is made up of individuals who stand in a variety of different relations to the agent in question, and who therefore have different degrees of interest and concern for her attitudes and conduct" (471). The question of when an agent is culpable or actually at fault, by contrast, does not allow for this sort of variation and fragmentation in our answer. Smith takes for granted that, with respect to the question of whether the moral agent is or is not responsible, we must secure some unequivocal answer that is not available to us when we are considering the stance of the moral judge who must decide if active blame is called for or appropriate.

Wallace (1994) certainly provides some resources for a reply to Smith's line of criticism. The first point to be mentioned is that, on Wallace's account, "active (p. 213) blaming" not only need not involve doing something to express blame, it may not even require feeling or engaging our emotions at all. As already noted, it is part of Wallace's "disjunctive formulation" of holding someone to reactive emotions to allow this to include simply believing that it would be appropriate to feel these emotions (23). Obviously, this qualification significantly closes the gap between what Smith describes as judgments about responsibility and actively blaming a person. At the same time, however, the connection between conditions of responsibility and holding a person to reactive attitudes must remain, for Wallace, because without this, judgments about responsibility or blameworthiness would be "rendered superficial" or "shallow." The aspect of "depth," which is essential to understanding what it is to be responsible, can be fully and completely appreciated only if we retain (some) reference to the attitudinal features found in the stance of the *judge* (51, 77–83). The force of Wallace's normative interpretation, insofar as it insists on retaining this connection between being and holding responsible, is that any analysis that severs this connection, as Smith would have us do, leaves our understanding of what it is to be

responsible incomplete and one-sided—lacking the needed and necessary psychological linkage between agent and judge.

Several features of the position that Smith takes on this issue are problematic. If we accept that conditions of being responsible and active blaming are to be distinguished in the manner Smith suggests then the following scenario would be entirely conceivable. We could find ourselves in a world where there are beings who are judged to be culpable, responsible agents but also no people who can be appropriately (actively) blamed. This would be a world in which blame had no place, even though it is populated by agents who are routinely judged to be responsible for their acts. It is not obvious that a world entirely drained of blame in this manner is one in which we could make adequate sense of responsible agency, or that we would be entitled to conclude that responsible agency was truly preserved. That is to say, in a world of this kind it is not obvious that the agents in question are really regarded as fully responsible for their actions. In Wallace's language, we may say a blameless world of this sort would be one in which judgments of responsibility lacked any “depth” or “force.” Any account of responsibility given in these terms is, to this extent, itself incomplete and insubstantial. By severing our assessments of culpability and fault from their (natural) connections and associations with conditions of (active) blame we erode the very fabric of moral life, and strip away the evaluative significance and motivational traction of moral judgment.

In responding to Smith's criticisms of Wallace, I have suggested that Wallace's discussion provides us with some relevant materials for dealing with Smith's general objection. There is, however, another way of approaching the question of the relationship between being and holding responsible that indicates that Wallace's (related) split between Strawsonian and Kantian components runs into difficulties and problems that are similar to those that Smith's views encounter. Specifically, Wallace's hybrid model, lends itself to the theoretical possibility of a moral world where a gap (i.e., an asymmetry) opens up between those who are responsible agents and those who can hold agents responsible. Consider Wallace's example of (p. 214) Mr. Spock (of *Star Trek* fame) who, as Wallace describes him, is not susceptible of human emotion and is, consequently, incapable of reactive attitudes and or of holding people responsible (Wallace 1994, 78n41). There is no reason, in principle, given Wallace's split between the Strawsonian and Kantian components of his analysis of responsibility, why an agent such as Mr. Spock may not be capable of “reflective self-control” (i.e., he is plainly “normatively competent” by Wallace's standards). At the same time, Mr. Spock is also, evidently, constitutionally incapable of holding himself or others responsible, because he lacks all capacity for reactive attitudes. For Wallace, there is no necessary or required connection between responsible agency and a capacity to feel or entertain reactive attitudes (i.e., between being a moral agent and being able hold oneself and others responsible). A world populated entirely by Mr. Spocks, such as the planet Vulcan (where Spock comes from), would be a world similar in kind to the world we have already envisioned when we considered Smith's views on the distinction between being responsible and active blame. A Vulcan world would be one in which, in Wallace's analysis, responsible agency (i.e., normative competence) would exist in circumstances where the responsible agents (i.e., the Vulcans) lack any capacity to hold

agents responsible. Because there is, according to Wallace's analysis, no necessary connection between a capacity to hold agents responsible (i.e., by means of reactive attitudes) and responsible agency itself, this is, on his account, at least a coherent and conceivable possibility. What is significant about the Vulcan world, as described, is that it is indeed an imaginary world taken from science fiction, quite unlike any real, recognizable human world with moral life as we know it.[2]

In a Vulcan world, as I have described it, responsible agency operates effectively and unimpaired in the complete absence of any capacity for reactive attitudes or moral sentiments. It is, however, highly questionable if our moral capacities, as we actually find them, would be undamaged or fully effective without a capacity to (actively) hold ourselves and others responsible. In the complete absence of any capacity to see ourselves and others as objects of reactive attitudes our capacity for recognizing and responding to moral considerations would surely be impaired. One good reason for supposing this to be true is that our relevant moral emotions give salience and significance to moral considerations and reasons. In the complete absence of any such emotional capacity, judgments of responsibility and their connection with moral considerations would lack the force and weight that we attach to them (via this mechanism). Agents such as Mr. Spock, and other Vulcans, would have a shallow and thin appreciation of moral reasons. Nor would they be motivated to recognize and respond to these reasons in the same way as (normal) human beings. If these general observations are correct, then it follows that for an agent to be responsible she must have a general capacity to hold herself and others responsible. There is, therefore, an intimate relationship between being and holding responsible as this concerns moral capacity. Considered from this point of view, we have reason to be skeptical about the suggestion that there could be a world in which there are agents who are responsible but who are, nevertheless, incapable of holding themselves and others responsible (Russell 2004).

(p. 215) History, Skepticism, and Pessimism: Hard Incompatibilism and Critical Compatibillism

Gary Watson, in his influential reflections on Strawson's “Freedom and Resentment” (Watson 1987b), identifies the lack of a plausible theory of exempting conditions as a general failing in Strawson's contribution. We have already considered some features of this criticism and possible lines of reply, such as Wallace's sketch of our powers of reflective self-control or normative competence. There is, however, a more specific vein of criticism that Watson pursues that cuts deep to the heart of issues that divide compatibilists and incompatibilists. The central concern here is what Watson describes as “the historical dimension of the concept of responsibility” (281). In order to explain the nature of this problem, Watson describes in some detail the case of Robert Harris, who committed brutal murders in California in 1978. Watson presents a detailed description of the events of the murders themselves, with a view to generating a strong reactive (retributive) response in his readers. What was particularly disturbing about this case was

the evidence of sadism and the complete lack of remorse. At the same time, there was no evidence of insanity or incapacity of any relevant kind (i.e., as described). Watson then switches the reader's attention to the historical background, detailing the horrors and extreme brutality of Harris's own childhood and adolescence. We are then invited to see Harris as victim, rather than a victimizer (275). The result of this switch in our attention and focus is not, Watson suggests, that it directly exempts Harris, but that it generates "ambivalence" in our response to him—emotional conflict is the product of these reflections (275). Watson goes on to suggest that cases such as this lead us to the general conclusion that, in the final analysis, we are not responsible for ourselves, because we are not the ultimate originators of our deeds (281–82). The upshot of these observations is that historical reflections of this kind make clear that "our ordinary practices are not as unproblematic as Strawson supposes" (283; also Nagel 1980; G. Strawson 1986, ch. 5; 1994; but contrast McKenna 1998a; Nichols 2007a).

Although Watson's own discussion stops short of endorsing a skeptical position, the general trajectory of his argument leads firmly in this direction. These sort of skeptical concerns about history suggest that it may not suffice to provide the Strawsonian strategy with an account of "reflective self-control" (i.e., along the lines of Wallace's approach). The incompatibilist or skeptical challenge may be pressed harder here by means of examples of implantation and manipulation. Counterexamples of this sort have been put forward, in one form or another, many times (see, e.g., Taylor 1963, 45–46; Dennett 1984; Pereboom 2007a; and Pereboom's essay in this volume). Regarding the general strategies we are concerned with, the basic concern is that for any preferred compatibilist conception of moral capacity (e.g., some mode of reflective self-control) it is theoretically possible that an external manipulator could implant the preferred structure in the agent and covertly control his conduct (p. 216) by this means (for a detailed discussion of this sort of case, see Kane 1996, 64–69). The difficulty for any compatibilist account—including the Strawsonian strategy we are considering—is that they have no principled reason to conclude that these manipulated individuals are not responsible agents. Counterexamples of this kind, drawing on "historical" considerations, lead us back down a slippery slope into skepticism. This is not the context in which to try and address these specific difficulties and objections to the wider compatibilist project. However, suffice it to say, for now, that objections of this kind require compatibilists to look either for further historical conditions on responsibility (e.g., excluding agents with "abnormal" or "deviant" histories) or to provide some nonhistorical basis, consistent with compatibilist commitments, that can account for why manipulation and implantation (appear to) pose a threat to responsible agency. (For various strategies see, e.g., Fischer and Ravizza 1998, ch. 8; McKenna 2004; Russell 2010)

Let us grant, for the moment, that the skeptical challenge cannot be effectively repelled by the arguments and strategy advanced by Strawson and his followers, we are still left (qua skeptic) with a significant set of problems on Strawson's analysis. In recent years there have been several important efforts to deal with some of these issues relating to the question of whether skepticism about responsibility is, for human beings, livable and/or bearable (i.e., worth living). The general issue that we have to deal with here is how skep-

ticism about moral responsibility relates to Strawson's account of the “objective attitude” and the question of “pessimism.” It is Strawson's view, as we have noted, that skepticism about moral responsibility should be interpreted as the view that our reactive attitudes are never justified or appropriate and must be altogether abandoned or suspended. (Strawson, of course, does not accept that skepticism about contracausal freedom or libertarian metaphysics itself justifies skepticism about moral responsibility.) It is also Strawson's view that a skepticism about responsibility, so interpreted, is psychologically impossible and, if possible, would be unbearably bleak and inhuman. The first of this pair of claims is part of his (strong) naturalism and the second is a feature of what he takes to be the linkage between skepticism and pessimism on this issue. We have already noted that even those who endorse Strawson's strategy of understanding responsibility in terms of our reactive attitudes need not accept his strong naturalist claim that it is psychologically impossible for us to live without the reactive attitudes. Wallace, for example, argues that our commitment to moral reactive attitudes and the associated system of moral expectations may be a cultural feature—one that other human cultures may not share with us (Wallace 1994, 3–2, 38–40, 64–65). To this extent, Strawson's strong naturalism does not seem essential to the wider position that he advances.

What about his views concerning the relationship between skepticism and pessimism in this sphere? It is certainly true, generally speaking, that skepticism about moral responsibility is widely associated with a pessimistic view of the human predicament (i.e., to the extent that responsibility is denied). Among Strawson's followers, however, there is some disagreement about the relationship between skepticism and pessimism. Some share Strawson's view that a life without any reactive attitudes would indeed be hopelessly bleak and humanly “impoverished” (see, e.g., Wolf 1981; (p. 217) Bennett 2008; and compare Smilansky 2001). Others, including Wallace, take a different view. Wallace, as we have noted, emphasizes the point that other forms of moral emotion may exist in the absence of reactive attitudes and it is a mistake (pace Strawson, Bennett et al.) to expand the class of reactive attitudes to include a wider range of emotions (e.g., reciprocal love) that are unconnected with expectations (Wallace 1994, 27; but see also P. F. Strawson 1962, 79). From this perspective there is no obvious or necessary linkage between a life entirely devoid of reactive attitudes, properly delineated, and Strawson's bleak description of living exclusively from the “objective stance,” with the “human isolation” that this would imply (81).

Whereas Wallace is persuaded by Strawson's broad anti-skeptical strategy, others who are not have more directly challenged his effort to present skepticism about moral responsibility as implying a deeply bleak view about our predicament in such a world. Among those who have challenged the simple connection between skepticism about responsibility and pessimism, Derk Pereboom (1995, 2001, 2007a) has been especially influential (see also Honderich 2002a, ch. 10; Sommers 2007). In several different contributions Pereboom has argued that skeptical worries about the ultimate source of conduct and character cannot be convincingly addressed by either compatibilist or libertarian theories of freedom and, for this reason, moral responsibility (i.e., understood in terms of “basic desert”) cannot be rescued from the various skeptical arguments that discredit it

(Pereboom 2007a, 86, 119, 123). Although much of Pereboom's attention is devoted to these skeptical arguments, in support of his “hard incompatibilist” position, it is his efforts to vindicate some form of (qualified) optimism consistent with his skepticism that is relevant to our present concerns. Whereas on the orthodox view that Strawson describes, skepticism about moral responsibility implies that a wide range of concerns and values attached to responsibility would be eroded, if not erased, Pereboom argues that this slide into pessimism is (grossly) exaggerated and largely unfounded.

Pereboom (2007a, 116–18) discusses a wide range of features of human life that may be thought to be threatened by skepticism about moral responsibility, including our sense of self-worth and our having meaning and purpose in life. It is, however, Pereboom's effort to find room for personal relations and a robust emotional life, consistent with his “hard incompatibilism,” where he most clearly diverges from Strawson. Pereboom grants that “the objective attitude,” as Strawson describes it, would be bleak and depressing. He denies, nevertheless, that our emotional lives would be impoverished in the way that Strawson suggests if we embrace skepticism or hard incompatibilism. He argues, in the first place, that only some forms of reactive attitude would be threatened by skepticism about moral responsibility. There are, he says, reactive attitudes that either would “survive” or have “analogues” that would be “sufficient to sustain good [personal] relationships” (Pereboom 1995, 269; 2007, 119). Moreover, many of those that do not survive or have no “analogues,” we would be better off without (e.g., certain kinds of anger and resentment). With this general position in view, Pereboom runs through a variety of personal emotions, such as forgiveness, gratitude, mature love, regret, and forms of “moral sadness,” that would persist or even thrive in the face of skepticism about moral (p. 218) responsibility in the sense of “basic desert” (Pereboom 1995, 269–71; 2001, 199–207; 2007a, 118–22). (See Pereboom's essay in this volume for further discussion of all of these topics.) Granted these alternative modes of reactive attitudes and personal emotions can survive and persist in the manner that Pereboom suggests, then skepticism about moral responsibility can be presented as being a potential source of genuine *optimism*—not a dreaded “difficult truth” that we must face up to (see also Watson 1987a, 284–86; Sommers 2007).

The various responses to Strawson that we have reviewed have challenged the way in which he suggests that skepticism about responsibility implies pessimism of some significant kind (e.g., despair, anxiety) about the human predicament. This is certainly a view that Pereboom, Honderich, and Sommers, among others, have questioned. By way of conclusion, however, I would like to raise some questions and doubts in the opposite direction. Let us assume that some version of Strawson's and Wallace's project of vindicating moral responsibility in terms of holding agents responsible on the basis of reactive attitudes can be defended (subject to further refinements and elaboration). Where does this leave us with respect to the optimism/pessimism duality that Strawson has drawn our attention to? A seemingly natural corollary of the suggestion that skepticism implies pessimism is that anti-skepticism (i.e., leaving responsibility in its place) must vindicate optimism—the view that with respect to the issue of moral responsibility we have no basis for finding the human predicament “difficult” or “depressing.” Strawson's language—like the

language of most compatibilism in general—encourages this “sunny” view (for an especially optimistic version of compatibilism as triumphing over the “gloom-leaders” of pessimism, see Dennett 1984, 7). It is worth mentioning, therefore, that there is another view that may be taken on this issue, one that regards the general vindication of moral responsibility along Strawsonian lines as a basis for a (moderate) pessimism about the human predicament.

The view I am describing has itself two core components. The first is a compatibilist theory of moral responsibility that builds on the work of Strawson, Wallace, and others (i.e., subject to further refinements). Among the relevant points of disagreement that will arise on this side of things, is whether or not we need a “revisionary” account—which will, in turn, depend on what we take our “ordinary intuitions” to be on this subject (see, e.g., Vargas 2007 and Vargas's essay in this volume). On the other side, where this view clearly diverges from most orthodox forms of compatibilism, it is argued that incompatibilist worries and concerns about ultimacy and sourcehood are well-founded and cannot simply be dismissed as illusory, confused, or groundless (as is argued, for example, by Dennett 1984, ch. 1; for criticism of this, see Russell 2002a). At the same time, this view—let us call it “critical compatibilism”—does not accept the incompatibilist or skeptical view that these pessimistic concerns about the impossibility of ultimacy for human agents licenses skepticism about moral responsibility itself. On the contrary, the key contention of critical compatibilism, so described, is that pessimistic reflections about the impossibility of ultimate agency and sourcehood are rooted in the thought that it is because we are morally responsible agents that these reflections on the limitations (p. 219) of agency (rooted in human finitude) present themselves as especially “difficult” or “hard truths” to deal with and accept. With respect to the source of these pessimistic features of critical compatibilism, two concerns are particularly significant. They are that responsible agency persists and endures in face of both fatalism and moral luck (Russell 2000, 2002, 2008). The mistake of the incompatibilist and skeptic, from the perspective of critical compatibilism, is that it takes these features to discredit and undermine responsible agency, whereas it is the persistence of responsible agency in face of these conditions that is the real and appropriate basis for pessimistic concern. Likewise, it is the mistake of complacent (optimistic) compatibilism, to try and conceal or minimize these difficult and problematic truths about the human predicament from us.

Interpreted this way, critical compatibilism, in its key claims, takes a position that is the opposite of Pereboom's “hard incompatibilism.”[3] The hard incompatibilist is a skeptic about moral responsibility but denies that this has the bleak and depressing implications that Strawson and others have attributed to it. The critical compatibilist, by contrast, rejects skepticism about moral responsibility but insists, contrary to the complacent compatibilist, that genuine and legitimate sources of pessimistic concern survive in these circumstances and conditions. For the critical compatibilist, reflection on our human predicament with respect to agency and moral responsibility is not a comforting source or basis for complacent optimism. Defeating the skeptical threat with respect to moral re-

sponsibility still leaves us having to deal with the deeper issues relating to human finitude and our associated limitations in this sphere (Russell 2002a, 2008, n.d.).

Concluding Remarks

In this essay, my primary concern has been to explore and describe the significance of P. F. Strawson's attempt to rebut the skeptical challenge to moral responsibility. Strawson's strategy, as we have noted, tries to chart a middle course between what he takes to be "the panicky metaphysics of libertarianism," on one side, and myopic, utilitarian-oriented compatibilism on the other. The strategy that Strawson pursues is, in important respects, a return to the traditional insights of the moral sense school—most notably, the views of David Hume and Adam Smith. (On the Hume-Strawson relationship, see Russell 1995, ch. 5.) At the same time, Strawson's method of turning away from narrow issues of conceptual analysis relating to the "logic" of freedom, constitutes a genuine and radical break with the standard literature and debate that dominated much of the twentieth-century discussion. Whether one is persuaded by Strawson's general strategy in "Freedom and Resentment" or not, it is fair to say that all those who currently work in this area must find a way through or around the arguments and issues that he has presented us with. The framework of the debate now includes the skeptical/naturalist and optimist/pessimist dualisms (p. 220) that Strawson introduced as key elements of his analysis. All parties in this debate must now locate their own positions with reference to this framework and take a clear stand on the basic points and issues that Strawson's contribution has brought to the fore.

Notes:

(1.) Wallace argues that not all reactive attitudes are moral reactive attitudes. It is only those reactive attitudes that involve moral expectations (obligations) backed by moral reasons that constitute the distinct class of moral reactive attitudes. There may, for example, be expectations based on etiquette that are also associated with reactive attitudes but lack any specific moral content. Wallace also argues that there are moral emotions other than moral reactive attitudes, such us shame, gratitude and admiration. Emotions of this kind cannot, he claims, be linked with (moral) expectations and reactive attitudes (Wallace 1994, 35–38).

(2.) On Smith's analysis some symmetry between being and holding responsible is preserved, in these circumstances, so long as we assume that Vulcan agents can judge when moral criticism is appropriate or called for (i.e., as distinct from any form of "active blaming"). For Wallace, in contrast, we can continue to view the Vulcan agents as genuinely responsible only if there are some (human) agents who are in a position to hold them responsible. In the absence of any (human) agents with reactive attitudes there would be no moral judges and, hence, no (deep) moral responsibility. Clearly, the difficulties that Smith and Wallace run into here are related but different.

(3.) Both hard incompatibilism and critical compatibilism may be described as nonstandard views, in that they reject the simple skepticism-pessimism (or anti-skepticism-optimism) linkage, as is generally assumed in the relevant literature (e.g., P. F. Strawson 1962). One of the more interesting features of Pereboom's contributions is that he challenges this orthodoxy.

Paul Russell

Paul Russell is Professor in Philosophy at the University of British Columbia. He has held a number of visiting positions, including Stanford University, the University of Pittsburgh, and the University of North Carolina at Chapel Hill. His published work includes 'Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility' (Oxford University Press, 1995) and 'The Riddle of Hume's Treatise: Skepticism, Naturalism, and Irreligion' (Oxford University Press: 2008). In 2010 he was the Fowler Hamilton Visiting Fellow at Christ Church, Oxford.