

# FREE WILL AND MORAL SENTIMENTS

## Strawsonian Theories

*Paul Russell*

Over the past few centuries, the free will debate has largely turned on the question of whether or not the truth of the thesis of determinism is compatible with the relevant form of freedom that is required for moral responsibility. This way of approaching the free will problem was fundamentally challenged by P.F. Strawson in his hugely influential paper “Freedom and Resentment,” which was published in 1962. In this paper, Strawson pursues a line of argument that can be found in the work of several major figures in the “moral sense school,” such as David Hume and Adam Smith. The strategy Strawson employs is one that begins with a complex and subtle description of the attitudes and practices that are constitutive of moral responsibility as we observe it in human life. According to Strawson, both sides in this debate fail to identify the real foundations of moral responsibility, which rests with the fabric of our human emotional psychology. When we start from *inside* these natural, human commitments, Strawson maintains, we are better placed to generate a viable and pertinent theory of human freedom as it relates to the requirements of moral responsibility.

### I

In order to provide a clear exegesis of Strawson’s core strategy, a few preliminaries are required. Strawson’s way of categorizing the main parties involved in this dispute is in important respects unorthodox and in some ways misleading. He labels the relevant parties in terms of their metaphysical attitudes in respect of the implications of the thesis of determinism. The classical compatibilists are described as “Optimists,” since they reject the suggestion that the truth of determinism would systematically discredit and dislodge our commitment to the attitudes and practices associated with moral responsibility. Those incompatibilists who are libertarians, and hold that we *are* free and responsible but that this requires the falsity of determinism, are labelled “Pessimists” in Strawson’s schema. Libertarians are “Pessimists” because they believe that the (skeptical) implications of determinism would be bleak and depressing for us all. Those incompatibilists who maintain that, whether determinism is true or not, we lack the relevant sort of freedom needed for moral responsibility are labelled “Moral Skeptics.”

Although Moral Skeptics are not identified in terms of their metaphysical attitudes, they generally fall into the Pessimist camp more broadly conceived (indeed, unlike libertarians, they maintain that there is no escape from skepticism and any pessimistic implications it may carry). In important respects, Strawson rejects all these positions. At the same time, he also aims to reconcile Optimists and Pessimists by finding some measure of truth in both their positions. It is, above all, Strawson's concern to discredit the claims of skepticism—a view, he argues, that is neither justified nor liveable.

In the context in which Strawson first presented his theory, there was an impasse between the views of classical compatibilism, which relies on a “one-eyed utilitarianism,” and various forms of (neo-Kantian) libertarianism, which rests on the “panicky metaphysics” of “contra-causal freedom” (Strawson 1962: 81–3). Strawson has little sympathy with either of these views and suggests that we must start our investigations elsewhere if we want to secure some agreement on this subject. The right place to start, he suggests, is with

the very great importance that we attach to the attitudes and intentions towards us of 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.

(Strawson 1962: 66)

This observation brings us to the role of reactive attitudes, which serve as the foundation for Strawson's naturalistic account of moral responsibility. In the case of personal reactive attitudes, “we demand some degree of goodwill or regard on the part of those who stand in [various] relationships with us” (Strawson 1962: 67). For example, in situations where one person is offended or injured by another, it is natural or normal to feel resentment. A full understanding of how reactive attitudes operate, and how some considerations serve to “modify or mollify” them, can provide us with a more general understanding of how moral reactive attitudes, such as praise and blame, operate and what relevance, if any, the thesis of determinism has with respect to these matters.

It is at this juncture that Strawson turns to his theory of excuses and exemptions and applies it to the justification of reactive attitudes. There are, he suggests, two broad categories of considerations that alter or inhibit our reactive attitudes towards an agent who has done us some injury or harm. First, there are specific excusing considerations that indicate that the agent's will was not of a kind that displays malice or a lack of due care or concern either for ourselves or others. In cases such as ignorance, accidents, or physical force, and so on, we see that the injury caused does not reflect any malicious or uncaring intent on the agent's part. However unfortunate the injury may be, the agent's *quality of will* is unobjectionable (Strawson 1962: 68). Another category of consideration is that of exemptions. Exemptions turn on the claim that the agent in question is in some way an inappropriate target or object of reactive attitudes—not just in the specific case at hand but more generally. In cases of this kind, we view the agent as somehow “abnormal” or “immature” (deranged, neurotic, an infant, etc.). In these circumstances, we are required to drop the participant stance that would engage our reactive attitudes and must instead adopt an ‘objective attitude.’ When we adopt the objective attitude to another human being we see them as an object of social policy but not as someone to reason with (Strawson 1962: 69, 70).

Once we have identified the role of reactive attitudes (e.g., praise and blame) as constitutive of responsibility, and explained how these responses are or are not altered and modified in light of excusing and exempting considerations, then we are better positioned to answer the vexed question concerning the implications of determinism for these attitudes and the practices associated with them (e.g., punishment). Would the truth of determinism, Strawson asks, “lead to the decay or the repudiation of all such attitudes” (Strawson 1962: 70, 71)? Strawson answers this question firmly in the negative. In relation to this issue Strawson argues for two related but distinct points:

1. The truth of determinism does not serve to systematically discredit our reactive attitudes.
2. Even if we reached such a radical conclusion on the basis of theoretical or philosophical reasoning of some kind, it would still be *psychologically* impossible for us to simply abandon or jettison our commitment to these reactive attitudes.

Let us consider each of these claims in turn. The first claim depends on Strawson’s analysis of excuses and exemptions. Strawson considers what relevance accepting the truth of determinism would have for our participant attitudes (e.g., resentment). Nothing about the thesis of determinism implies that agents are always ignorant about what they are doing, nor that they never act intentionally or that everything that is done is an accident or inadvertent (Strawson 1962: 70, 71). Similarly, nothing about this thesis implies that every agent is in some relevant way “abnormal” or “immature” (Strawson 1962: 71). If this was the case, then we would have to adopt the objective attitude to everyone and drop all our reactive attitudes. However, so long as an agent is not incapacitated from ordinary personal relationships and is neither a child nor abnormal (e.g., mentally ill, etc.), then no such policy is required of us. The upshot of this is that, contrary to the Pessimist, the thesis of determinism has no radical skeptical implications for our commitment to the attitudes and practices involved in moral responsibility.

How convincing is this line of argument? Strawson’s argument, as presented, seems seriously incomplete, if not fatally flawed. The key issue here rests with the account offered of exemptions. Strawson’s remarks on this subject, although crucial to his argument, are very brief. More importantly, what he says tends to conflate issues of “abnormality” and “immaturity” with that of incapacity. In a key passage, Strawson dismisses any effort to *generalize* exemptions on the basis of worries about determinism and argues that “it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition” (Strawson 1962: 71). Although Strawson acknowledges that this might come across as “too facile,” his account of what the relevant moral capacities do involve is very thin—too thin to bear the weight of the ambitious anti-skeptical conclusion that it must support. He does indicate what the relevant moral capacities do *not* involve; namely any form of libertarian “contra-causal freedom”—a condition of responsibility that, he claims, “cannot be coherently described” (Strawson 1980: 265). Beyond this, however, his positive or constructive remarks are both brief and sketchy, providing just a few sprinkled remarks concerning the need for a sense of reality (Strawson 1962: 72); an ability to be reasoned with (Strawson 1962: 69, 70); moral sense or a susceptibility to reactive attitudes (Strawson 1962: 75–7); and an awareness of our conduct in terms of conscious purposes (Strawson 1962: 75, 76, 78). Some incompatibilists are skeptics but not libertarians (i.e., they are not Pessimists in the narrower sense) and they may well agree with Strawson that there is no available

coherent concept of freedom that serves the needs and aspirations of libertarians—but this does not defeat their skepticism. On the contrary, they will agree with (libertarian) Pessimists that the sort of capacities that Strawson briefly alludes to also fail to suffice as an account of the sort of moral capacities required to justify our reactive attitudes or moral sentiments. If something can be said on behalf of compatibilism with regard to this important matter, Strawson’s account fails to provide it. This leaves the door open to the skeptic, as well as those who believe they can provide a more robust form of agency on the basis of the metaphysics of indeterminism and contra-causal freedom. Given these difficulties, even those broadly sympathetic to Strawson’s approach must conclude that his argument falls short at this crucial juncture.

Strawson assumes that the skeptical challenge based on generalized exempting conditions can be swiftly swept aside on the ground that universal “abnormality” is impossible. We have noted that this way of interpreting and refuting the skeptical challenge is confused. Strawson has, however, another argument to back up his case against skepticism. Whatever our “theoretical” or skeptical philosophical reflections may suggest, Strawson argues, it is “practically inconceivable” for any reasoning of this kind to systematically dislodge our commitment to the reactive attitudes (Strawson 1962: 71, 72, 77; see also Strawson 1985: 11, 13, 39, 41).

The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any things as inter-personal relationships as we normally understand them . . . A sustained objectivity of interpersonal 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 were a theoretical ground for it.

(Strawson 1962: 71, 72; and also Strawson 1985: 39)

According to Strawson, then, our human nature inoculates us against all radical skeptical threats of the sort that animate and concern the Pessimist. Skepticism is taken to imply that we must entirely jettison all reactive attitudes, which would, he claims, result in us adopting the “objective attitude” at all times to all people. This is something, if put into practice, that would certainly reduce us to a bleak and dehumanizing condition (Strawson 1962: 71–3, 77, 79). In summary, skepticism and the objective attitude that it implies, is not only *unliveable*, it would be *unbearable*. If it were a matter of decision and choice—as it is not—we would have every reason of a practical kind to reject any proposed move to a universal objective standpoint, bereft of the web of personal and emotional responses that make us human.

Does this argument fare any better than Strawson’s previous argument to defeat skepticism? The answer to this is again negative. There are several different lines of possible criticism but I will focus on a particular confusion that lies at the heart of Strawson’s strong naturalist rejoinder to the Skeptic or Pessimist. With respect to the claim that we are in some sense naturally committed to reactive attitudes we need to distinguish two distinct claims. First, there is the claim that our human constitution naturally renders us liable or prone to various forms of emotion, including reactive emotions broadly conceived, and that this liability or dispositional capacity neither requires nor is capable of rational justification (Strawson 1962: 81). Although this claim is credible it does

not serve to discredit skepticism in the way Strawson supposes. Strawson needs a stronger form of naturalism, one that claims that we will continue to entertain and experience particular or individual *tokens* of reactive attitudes irrespective of any skeptical arguments or general theoretical truths that may be advanced against them. We might distinguish this stronger claim as token-naturalism as opposed to type-naturalism of the weaker kind, as associated with our emotional propensities and dispositions. (It is, of course, the distinction rather than the labels that matter here.)

Clearly, skeptical arguments could systematically discredit and dislodge our commitment to all *tokens* of reactive attitudes, even if we retain the weaker, *type-naturalist* propensity to these emotions. Imagine, for example, that we discover that everyone is subject to social manipulation or concealed control and conditioning of some kind. In these circumstances, the objection runs, the fact that we may retain a *liability* to reactive attitudes does not show that *any* tokens of this kind would still be justified. We may well find circumstances of this kind incredible or unlikely but they are, nevertheless, entirely coherent and conceivable and this reflects a basic distinction between two distinct levels of concern arising from the skeptical challenge and the associated demand for justifications. Contrary to Strawson's presentation of his argument, the skeptical objection is best interpreted at the *token* level, where it is not our liability to reactive attitudes that is in question but whether any tokens of these reactive attitudes can ever be justified as they arise (i.e., systematically, on a case by case basis).

Strawson and his followers might try to tough-out this approach by insisting that we should accept *both* type and token naturalism but this view has little to be said for it. First, it is simply not credible that we are psychologically constrained to retain a commitment to reactive attitudes that we reflectively judge to be unjustified or inappropriate. Second, even if this were true, it would not serve to discredit or refute skepticism. If our human psychology commits us to entertaining and experiencing reactive attitudes in these circumstances this may well be *disturbing* but it does not show that the token reactive attitudes in question are appropriate or justified, which remains the relevant focus of concern for the skeptic.

It is Strawson's further contention that were the case for skepticism accepted and followed through in practice (i.e., we undertake, systematically, to stop entertaining reactive attitudes), then a bleak inhuman condition involving the universal adoption of the objective attitude must follow. This further line of argument has also been challenged. Several critics have argued that Strawson's taxonomy of the reactive attitudes and the way they are related to moral and personal emotions, more generally, overstates their scope and extent (See, e.g., Wallace 1994: esp. Chapter 2; Pereboom 2001: esp. 199–207; Sommers 2007). One of the features of Strawson's analysis that may be questioned here is his assumption that the reactive attitudes comprehend all the various emotions involved in personal relationships (e.g., love and friendship). If these are distinct and separate categories of emotion then the whole-scale jettisoning of reactive attitudes need not imply the cold, bleak universal stance of the objective attitude that Strawson foresees. Moreover, while personal relations may be humanly inescapable, it does not follow that reactive attitudes are similarly inescapable. There may, moreover, be some forms of reactive attitude that are not threatened by the skeptical challenge, just as there may be moral emotions and responses that should not be assimilated to (moral) reactive attitudes.

All these matters leave room for modifications and amendments to Strawson's analysis of the implications of the skeptical challenge. They also leave room for more

detailed and elaborate replies on behalf of Strawson (see, e.g., Shabo 2012). Without attempting to settle the issue here, it is safe to say that both Strawson's critics and his followers would generally agree that his account of the relationship between skepticism, objectivity and pessimism calls for a more nuanced and fine-grained analysis of reactive attitudes and what would be involved in attempting to live a human life altogether without them.

## II

We have noted that one of the most significant gaps in Strawson's approach to moral responsibility is his thin account of moral capacity—a gap that leaves his argument open to the Pessimist objection that we still require a more robust conception of moral freedom (i.e., some mode of contra-causal freedom). Although this gap is problematic, several more recent compatibilist contributions have suggested ways of closing it, consistent with the compatibilist requirements, in general, and the Strawsonian approach, in particular. The most influential of these has been rational self-control or reason-responsive theories. A theory of this kind has been advanced by R. Jay Wallace, who weaves it into a broadly Strawsonian framework (Wallace 1994; see also Fischer and Ravizza 1998, who present a related account of reasons-responsiveness). More specifically, Wallace combines a Strawsonian account of holding responsible with a Kantian theory of moral agency. To understand the relevant conditions of moral responsibility, Wallace claims, we need to provide an account of when it is *fair* to adopt the stance of holding an agent responsible. His account of holding responsible has, in this sense, priority over his account of being responsible. (See the discussion further below.)

Wallace's Strawsonian account of holding responsible involves a 'narrower' and more restricted interpretation of moral reactive attitudes. To hold an agent morally responsible, on this view, is to hold the person to moral expectations one accepts (Wallace 1994: 51). Our moral expectations are themselves supported by moral reasons that serve to generate obligations (Wallace 1994: 63–4). To hold someone to expectations of this kind must itself be understood in terms of our susceptibility to (moral) reactive attitudes (Wallace 1994: 21). When an agent violates a moral obligation a reactive attitude (e.g., resentment, indignation, or guilt) is called for and appropriate. There is, therefore, a mutual dependence between moral expectations and moral reactive attitudes. Although this 'normative' theory provides a tidy account of the structure of the beliefs involved in holding an agent responsible it also considerably 'narrows' the scope of moral responsibility. That is to say, given that reactive attitudes are appropriately triggered only when a moral obligation has been voluntarily violated, it follows that reactive attitudes are always negative in nature (Wallace 1994: 63, 64, 71). If this is the case then, on this analysis, an asymmetry arises with respect to responsibility, since there is no room left for holding agents responsible in terms of 'positive' reactive attitudes (e.g., approval, praise, etc.). Any account of this kind, it may be argued, is not only one-sided, it presents a truncated and impoverished description of moral life as it relates to responsibility (Russell 2013).

A further difficulty of Wallace's narrow account of moral responsibility is the way in which it commits him to the general apparatus of what Bernard Williams has described as "the morality system" (Williams 1985: Chapter 10; and cp. Wallace 1994: 39, 40, 64–6). The morality system is understood as a particular form of ethical life, associated with our modern Western culture. The narrow view, therefore, is committed to an

account of moral responsibility that has a ‘local’ bias toward our modern Western culture, as manifest in the morality system. This bias excludes other cultures, such as the ancient Greeks or shame cultures, as being, at best, ‘analogous’ to ours, since they involve different moral beliefs with distinct patterns of emotional response (Wallace 1994: 65, 66). Quite apart from anything else, this form of ‘localism’ or bias to the morality system, considered as serving as the relevant standard of true or real responsibility, removes all commitment not just to token-naturalism but also to type-naturalism. It discredits (Strawsonian) type-naturalism because it maintains that the framework of moral responsibility is erected around culturally local forms of moral emotion. If this is correct, then, contrary to Strawson, we still require an external rational justification for this entire framework and remain vulnerable to global skepticism at this level (Russell 2013).

Let me now turn to the other strand in Wallace’s Strawsonian program, his Kantian theory of moral agency, understood as an effort to provide a theory of moral capacity that does not presuppose any “strong freedom of the will” and alternative possibilities (Wallace 1994: 86). Wallace’s theory suggests, instead, a requirement of “normative competence.” A theory of this kind, Wallace argues, should be understood in terms of (i) the power to grasp and apply moral reasons, and (ii) the power to control or regulate behavior in light of such reasons (Wallace 1994: 86, 157). Agents who have these powers are capable of “reflective self-control.” Although determinism may deprive us of genuine alternatives, it does not in itself deprive us of the relevant paired powers of normative competence that Wallace has identified. In embracing this general picture of moral capacity and normative competence, Wallace joins the ranks of a number of contemporary compatibilists who have argued that our dispositional abilities of rational self-control serve as the relevant basis of freedom and moral responsibility (Dennett 1984; Scanlon 1998; Wolf 1990; and Fischer and Ravizza 1998). What is distinctive about Wallace’s approach is the way he has fused this theory with a Strawsonian understanding of holding responsible (but see also McKenna 2012 and Shoemaker 2015).

Although Wallace’s Kantian theory of agency fills a significant gap in Strawson’s overall naturalistic argument, it encounters its own set of difficulties and objections. The most important of these is the question of whether his account of rational self-control suffices as the relevant base capacity for responsible agency. The obvious objection, from an incompatibilist perspective, is that the mere possession of such (dispositional) powers does not give the agent control over the way in which they are actually *exercised*. What the incompatibilist is looking for here, and will not find in Wallace’s discussion, is a convincing account of how it can be *fair* to hold a person responsible for conduct that flows from powers that are exercised in ways over which the agent has no (ultimate or final) control. The deep worry here, about fairness, is itself rooted in concerns about history and luck.

A particularly influential discussion of this problematic dimension of the Strawsonian strategy is presented by Gary Watson (Watson 1987). One of Watson’s central concerns is “the historical dimension” of the problem of responsibility, which he develops by describing in some detail an actual case of a vicious murderer (who was eventually executed). Watson describes two distinct perspectives from which we may consider this case (and other cases similar to it). The first aims to trigger our strong negative reactive attitudes by focusing on the brutal and cruel nature of the crime itself. With respect to this crime there were no evident excusing or exempting considerations other than the “extreme evil” of the act itself (Watson 1987; 94–7). From another perspective, however, we may focus our attention on the background formative conditions on the

character and motivations of the criminal, which in this, as in many other cases, involves a brutal and harsh childhood and adolescence. A “history that involves unfortunate formative circumstances” (Strawson 1962: 70; Watson 1987: 90) does much to explain “the roots of evil” and why the criminal ended up doing what he did. Although reflections of this kind may not, necessarily, generate skepticism, they will (or should) generate some degree of “ambivalence” (Watson 1987: 101). We oscillate between viewing the criminal as a victimizer and a victim—an instability in our responses that generates deep emotional conflict.

What is operating and affecting our responses and reactions in cases of this kind is an awareness of moral luck. Watson put the problem this way:

If determinism is true, then evil is a joint product of nature and nurture. If so, the difference between any evil person and oneself would seem to be a matter of moral luck. For determinism seems to entail that if one had been subjected to the internal and external conditions of some evil person, then one would have been evil as well. If that is so, then the reflections about moral luck seem to entail that the acceptance of determinism should affect our reactive attitudes . . . determinism seems to be relevant to reactive attitudes after all.

(Watson 1987: 103, 104)

The only way to avoid this is to employ the apparatus of libertarian metaphysics in an effort to show how we can be responsible for ourselves. According to Watson, this is an “unbearable burden” and a hopeless enterprise (Watson 1987: 106). The underlying intuition supporting these libertarian ambitions is that we must somehow have consented to evil and unless the consent was undetermined we would not truly be originators of our deeds—we would be mere products and not producers (Watson 1987: 107).

These reflections about “the historical dimension” of the concept of responsibility take us well down the path to skepticism. However, although Watson insists that the historical dimension is “a potential source of skepticism” in relation to our attitudes and practices of holding responsible, he resists skepticism. Instead, he considers the possibility that we might strip away “retributive sentiments” but still retain a commitment to holding responsible in the more limited sense of making an appeal or demand to others as moral agents, without simply collapsing into the objective attitude (Watson 1987: 110, 111). A number of others, pursuing similar lines of thought, have argued that the implications of historical considerations take us well down the road to full-blown skepticism about moral responsibility (see, e.g., Strawson 1994; Pereboom 2001; Levy 2011).

### III

Another line of criticism levelled against the Strawsonian strategy is that while accounts of moral capacity in terms of our powers of rational self-control are the right basis for responding to the skeptical challenge, a solution along these lines serves to show that the Strawsonian approach, with its focus on reactive attitudes and holding responsible, is the wrong foundation for a satisfactory theory of responsibility. The general problem here, as Angela Smith argues, is that the Strawsonian strategy turns on a confusion between conditions under which it is appropriate to judge someone to *be* responsible with conditions under which it is fair to blame them or *hold* them responsible (Smith 2007). Being responsible is a matter of the agent being ‘culpable’ or ‘at fault’—as when



an agent fails to act on available moral reasons. Blaming, however, involves something more than the mere judgment that the agent was at fault. Although it may not involve overt or expressed retributive activity, it must involve some relevant degree of feeling, whereby the agent is resented or the object of indignation or anger. Although blaming may presuppose fault or culpability, it does not follow from this that blame is always appropriate when an agent is judged culpable or at fault. Even if the agent is clearly responsible for wrongful conduct of some kind, various considerations, distinct from the considerations relating to the agent being responsible, may prohibit or discredit (active) blaming in these circumstances.

The criticisms offered above rely on the supposition that being responsible is both prior to and independent of holding responsible. The problem with the Strawsonian approach, critics argue, is that it gets this relationship backwards, since Strawson's analysis *begins* with an account of the reactive attitudes. There are, however, several features of this account that may be challenged. In order to understand the relationship between being and holding responsible, let us return to Wallace's split between Kantian agency and Strawsonian holding responsible. On Wallace's construal, it is entirely probable that an individual may have powers of rational self-control but lack any capacity for reactive attitudes. In these circumstances, an agent would be responsible but unable to understand (i.e., from the 'inside') what is involved in being held responsible or holding others responsible. One important objection to this is that, with regard to actual (real) human beings, there is an intimate relationship between the effective development and functioning of responsiveness to moral reasons and the possession of a moral sense, where the latter involves an ability to feel and understand moral sentiments and reactive attitudes. We generally find that agents who satisfy conditions of reasons-responsiveness also satisfy the condition of moral sense and this is not accidental. The two capacities are interdependent and interconnected. Responsible agents, therefore, must also be capable of holding themselves and others responsible—we cannot separate these issues in the way that some of Strawson's critics suppose (Russell 2004; McKenna 2012).

Michael McKenna has recently advanced a new theory of responsibility that is broadly Strawsonian in nature and that pays particular attention to the relationship between being and holding responsible (McKenna 2012). Although McKenna agrees with Strawson that we cannot understand responsibility without reference to holding responsible, and that holding responsible must be explained in terms of the operation of reactive attitudes, he argues that we need to find a more balanced account of the relationship between being and holding responsible—an account that rejects the suggestion that either one is in some way prior to the other (McKenna 2012: 3, 4). The key to understanding this relationship, and the symmetry that it involves, McKenna suggests, is to consider it in terms of the analogy between responsibility and conversation. Clearly a competent speaker must not only be able to express herself but also be able to interpret and understand those who may reply to what she says. This is essential if the speaker is to be able to properly appreciate the significance of her own remarks and contribution to a conversation. In the same way, McKenna argues, a responsible agent must, like a competent speaker, be able to appreciate the significance of her act and the quality of her will with which she acts. This is possible only if she is able to interpret the way her acts are received and responded to by others. It follows that competence as a moral agent requires abilities on both sides of this divide and these skills cannot be prised apart (McKenna 2012: 86, 97, 99, 100, 196, 213).

The conversational analogy suggests a “three-stage” structure for our understanding of responsibility. The first stage is “moral contribution,” which occurs when an agent “opens up the possibility of a conversation about the moral value of her action” (McKenna 2012: 88, 89). The second stage is “moral address,” when members of the moral community hold the agent morally responsible by directing their reactive attitudes at her (e.g., by blaming her). The third and final stage is “moral account,” when the agent has the opportunity to justify, excuse or repudiate her conduct. As the analogy with conversation or “dialogue” suggests, no priority should be given to the stage of moral address (McKenna 2012: 212, 213). At the same time, as McKenna points out, it is also clear that no agent who is incapable of making sense of the reactive emotions of others in response to her conduct can be fully capable of appreciating the quality of will with which she acts or the significance of what she does.

McKenna employs the conversational model to elucidate our understanding of blame and desert. The basic point of blame is to communicate with the wrongdoer and, as such, blame is both public and directed at the wrongdoer. Given these conversational constraints, blame will generally involve some measure of harm to the agent (McKenna 2012: 135, 153). The value of blame, according to this account, rests with several non-instrumental goods that it secures. These include the good that is manifest in the blamer’s commitment to morality and the generation of “dialogue aimed at resolution and reconciliation” (McKenna 2012: 167–9). What is crucial, however, as the conversational model suggests, is that blaming responses directed at the agent are *fitting*, where this is understood in terms of modes of expression that serve to “move the dialogue further along” (McKenna 2012: 142).

Another important contribution to Strawsonian theory that has appeared recently is David Shoemaker’s “tripartite theory of responsibility” (Shoemaker 2015). In *Responsibility from the Margins*, Shoemaker aims to extend and improve upon Strawson’s project to provide an account of moral responsibility based upon a description of the role of moral sentiments in this sphere. As with McKenna’s study, Shoemaker also pays particular attention to the importance of the agent’s quality of will and the way in which this relates to the full range of reactions that are elicited. According to Shoemaker, however, there is a significant failing in most quality of will theories—including Strawson’s. These theories tend to be “monistic,” by which Shoemaker means that they fail to adequately distinguish the various kinds of quality of will, along with the distinct forms of sentimental response that they occasion. As a result of this coarse-grained approach, Shoemaker argues, we have difficulty dealing with “marginal agents” (e.g., psychopaths, autistics, those who suffer from dementia, etc.) who do not fit neatly into the framework provided by monistic theories. Shoemaker maintains that his “tripartite theory” offers a more nuanced account of these matters, one that deals more adequately with the complexities that these marginal cases present to us.

There are, Shoemaker suggests, three distinct types of responsibility that correspond to three aspects of quality of will. The “three agential factors” that we care about are character, judgment and regard. Our responsibility sentiments, he argues, have “fulfillment conditions” that depend on what specific sort of quality is being targeted. For example, quality of regard is evaluated in terms of anger or gratitude, whereas quality of character is captured by emotional responses such as disdain and admiration. Each type of responsibility correlates with the relevant feature or dimension of quality of will and the distinct sort of responsibility sentiments that are suitable to them. Attributability responsibility is concerned with “volitional structures” and the way they reflect our deep

selves. Answerability responsibility focuses on the judgments agents make about the worth and relative weight of reasons and distinct responses involved and, so considered, should be construed as demands for justifications. Finally, accountability responsibility involves the distinct responses of anger or gratitude in relation to the degree of regard and empathy that an agent may display. The particular value of displaying anger, Shoemaker maintains, is to be found in the role that it plays in preserving the moral community and generating empathetic concern.

Shoemaker employs the structure of his pluralist, tripartite theory to help us make better sense of the various “hard” cases we find on the margins of the moral community. With the tripartite apparatus in place we may find that there are agents who are responsible in all three aspects and others who are responsible in none. Just to take one important example, in the case of psychopaths their capacity for regard is damaged due to their incapacity for empathy, and so they are not accountability responsible. However, psychopaths may still manifest quality of will in respect of character and be capable of some form of rational judgment. To this extent, they may be considered responsible in the attributability and answerability senses but not the accountability sense. Shoemaker provides similar, nuanced analyses of other marginal cases and draws significant practical conclusions about how we should treat and approach these individuals.

Let us now return to a theme in Strawson’s contribution that we considered earlier—the question concerning the relationship between skepticism and pessimism. As we noted, Strawson’s remarks on this subject suggest that skepticism about moral responsibility implies a universal objective attitude and that we would find a life of this nature to be bleak and humanly impoverished (Strawson 1962: 71, 73, 79). This is a claim that both Strawson’s followers and critics have challenged (e.g., Wallace 1994: 27, 28; Pereboom 2001: 90–8). The question arises, nevertheless, as to whether or not the *defeat* of skepticism, along the naturalist lines that Strawson advances, would serve to vindicate *optimism*? Strawson’s own remarks certainly suggest the more optimistic view. However, against this suggestion, it may be argued that nothing that Strawson or his followers have argued or advanced serves to discredit the view that agents who satisfy the general conditions of freedom and responsibility, as described, may still be subject to various modes of fate and luck. Although the *critical* compatibilist may agree that this does not serve to support *skeptical* conclusions, the persistence of responsible agency in these circumstances can hardly be construed as comforting or without its own disconcerting and disturbing implications. If this is correct, then the success of the Strawsonian strategy, while it may serve to defeat skepticism, does not serve to vindicate any form of easy or complacent optimism (Russell, forthcoming).

#### IV

It is evident that Strawson’s approach to the free will problem has generated an extensive and rich set of replies and responses over the past half-century. Moreover, as we have noted, Strawson’s contribution has changed in fundamental terms much of the way in which this debate is conducted. The key features that are essential to this approach include the following. First, from a methodological point of view, the Strawsonian approach places emphasis on the relevance of descriptive moral psychology for our understanding of these matters. Second, and related to this, the Strawsonian approach identifies moral sentiments or reactive attitudes as the foundation of any

adequate analysis of the conditions of responsibility and the forms of freedom that this is required for it. Third, the Strawsonian approach employs these broadly naturalistic components (i.e., the two items previously mentioned) to defeat the skeptical threat as associated with the free will problem. Finally, this approach may be presented as aiming to vindicate common sense against various forms of philosophical extravagance rooted in an excessive rationalism. The deep diagnosis provided is that the free will problem is generated by misplaced philosophical demands for modes of justification that have their roots in a failure to ground our investigations in an account of the operations of human nature as we observe them. Each one of these components faces objections and difficulties of their own. Nevertheless, taken as a whole, the theory provided is both ambitious and sophisticated and one that all parties to this debate must address and consider.

## Bibliography

- 
- Dennett, D. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*. Oxford: Clarendon Press.
- Fischer, J. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Levy, N. (2011) *Hard Luck*. Oxford: Oxford University Press.
- McKenna, M. (2012) *Conversation and Responsibility*. New York: Oxford University Press.
- McKenna, M. and Russell, P. (eds) (2008). *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment."* London: Ashgate.
- Pereboom, D. (2001) *Living Without Free Will*. Cambridge: Cambridge University Press.
- Russell, P. (1992) "Strawson's Way of Naturalizing Responsibility," *Ethics* 102: 287–302.
- Russell, P. (1995) *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility*. New York: Oxford University Press.
- Russell, P. (2004) "Responsibility and the Condition of Moral Sense," *Philosophical Topics* 32: 287–305.
- Russell, P. (2011) "Moral Sense and the Foundations of Responsibility," in R. Kane (ed.), *The Oxford Handbook of Free Will*, 2nd edn. New York: Oxford University Press, pp. 199–220.
- Russell, P. (2013) "Responsibility, Naturalism and 'The Morality System,'" in D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*, Vol. I. New York: Oxford University Press, pp. 184–204.
- Russell, P. (forthcoming) "Free Will Pessimism," in D. Shoemaker (ed.), *Under Consideration with Oxford Studies in Agency and Responsibility*, Vol. III. New York: Oxford University Press.
- Russell, P. and Deery, O. (eds) (2013) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press.
- Scanlon, T.M. (1998) *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Shabo, S. (2012) "Where Love and Resentment Meet: Strawson's Intrapersonal Defence of Compatibilism," *The Philosophical Review* 121: 95–124.
- Shoemaker, D. (2015) *Responsibility from the Margins*. New York: Oxford University Press.
- Smith, A. (2007) "On being responsible and holding responsible." *Journal of Ethics* 11: 465–84.
- Sommers, T. (2007) "The Objective Attitude," *The Philosophical Quarterly* 57: 321–41
- Strawson, G. (1994) "The Impossibility of Ultimate Moral Responsibility," Reprinted (with some modifications) in Russell, P. and Deery, O. (eds) (2013) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press, pp. 363–78.
- Strawson, P.F. (1962) "Freedom and Resentment," Reprinted in Russell, P. and Deery, O. (eds) (2013) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press, pp. 63–83.
- Strawson, P.F. (1980). "Replies," in Z. Van Straaten (ed.), *Philosophical Subjects: Essays Presented to P.F. Strawson*. Oxford: Clarendon Press, pp. 260–96.
- Strawson, P.F. (1985) *Skepticism and Naturalism: Some Varieties*. London: Methuen.
- Watson, G. (1987) "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," Reprinted in Russell, P. and Deery, O. (eds) (2013) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press, pp. 84–113.
- Wallace, R. Jay (1994) *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Williams, B. (1985) *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press.
- Wolf, S. (1990) *Freedom within Reason*. New York: Oxford University Press.

## Further Reading

---

- McKenna, M. (2012) *Conversation and Responsibility*. New York: Oxford University Press. (An important effort to develop the Strawsonian theory by way of drawing on the analogy between being capable of participating in a conversation and being a member of the moral community and the modes of communication that this involves.)
- McKenna, M. and Russell, P. (eds) (2008). *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's "Freedom and Resentment."* London: Ashgate. (A collection of some of the most important and influential discussions in response to P.F. Strawson's "Freedom and Resentment".)
- Russell, P. (1992) "Strawson's Way of Naturalizing Responsibility," *Ethics* 102: 287–302. (An analysis and critique of the core arguments advanced in Strawson's "Freedom and Resentment" (reprinted in McKenna and Russell 2008).)
- Shoemaker, D. (2015) *Responsibility from the Margins*. New York: Oxford University Press. (An ambitious and wide ranging application of Strawsonian themes arguing for the 'pluralist' thesis that there are three distinct modes of responsibility that need to be distinguished, which helps us make better sense of a number of difficult cases arising on the fringes of the moral community.)
- Strawson, P.F. (1962) "Freedom and Resentment," Reprinted in Russell, P. and Deery, O. (eds) (2013) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press, pp. 63–83. (The seminal statement of the Strawsonian approach, around which most of the subsequent philosophical literature has developed.)
- Wallace, R. Jay (1994) *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press. (A highly influential effort to revise and reconstruct a broadly Strawsonian account of holding responsible with a Kantian theory of moral agency.)
- Watson, G. (1987) "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," Reprinted in Russell, P. and Deery, O. (eds) (2013) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. New York: Oxford University Press, pp. 84–113. (An illuminating and sympathetic examination of problematic aspects of Strawson's theory, addressing in particular issues of luck and history as they generate skeptical concerns.)

## Related Topics

---

Reasons-Responsive Views  
 Classical Compatibilism  
 Skeptical Views about Free Will  
 Nonstandard Views  
 David Hume  
 Blame