

## Optimism about Moral Responsibility

Jacob Barrett

Abstract: In his classic “Freedom and Resentment,” P. F. Strawson introduces us to an “optimist” who believes that our moral responsibility practices are justified by their beneficial consequences. Although many see Strawson as a staunch critic of this consequentialist position, he actually claims only that there is a gap in the optimist’s story, and that the reactive attitudes are needed to fill it. In this paper, I fill in this gap. I show how optimism can be suitably modified to reflect an appreciation of the reactive attitudes, and argue that the ensuing position—on which our moral responsibility practices, taken as a whole, are justified both by their regulation of behavior in socially desirable ways and by their enabling of valuable interpersonal relationships—provides us not only with a plausible justification of our moral responsibility practices, but also with a fruitful framework for criticizing particular features of these practices and for evaluating potential reforms.

### I.

In “Freedom and Resentment,” P. F. Strawson sets up a debate between a “pessimist” and an “optimist.” The pessimist is an incompatibilist. He argues that the truth of determinism would render us unfree, and that we cannot be morally responsible if we are not free. The optimist is a compatibilist. He argues that our practices of holding one another morally responsible are justified by their beneficial consequences—in particular, by their regulation of behavior in socially desirable ways—and that this depends not at all on the truth or falsity of determinism.

Strawson rejects the pessimist’s incompatibilism, but expresses dissatisfaction with the optimist as well. He complains that there is a “lacuna in the optimistic story” (FR 4), that “to speak in terms of social utility alone is to leave out something vital in our conception of

these practices” (FR 24).<sup>1</sup> The optimist, says Strawson, is a “one-eyed utilitarian,” because “[h]e seeks to find an adequate basis for certain social practices in calculated consequences, and loses sight... of the human attitudes of which these practices are, in part, the expression” (FR 25). In other words, and in terms that are now familiar, Strawson’s charge is that the optimist “neglects or misconstrues” the *reactive attitudes* we adopt when holding one another morally responsible (FR 24). Some of these attitudes, including “gratitude, resentment, forgiveness, love, and hurt feelings,” Strawson labels the “personal reactive attitudes” (FR 5). These are not attitudes we calculatingly take up in order to affect behavior, but “reactions to the quality of others’ wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern” (FR 15). Strawson explains:

The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves; or at least on the expectation of, and demand for, an absence of the manifestation of active ill will or indifferent disregard (FR 15).

The other reactive attitudes also reflect an expectation of, and demand for, good will, but they have different targets. Moral indignation and disapprobation, for example, are “vicarious analogues” of the personal reactive attitudes that are “reactions to the qualities of others’ wills, not towards ourselves, but towards others” (FR 15). And attitudes like guilt, remorse, and shame, are “self-reactive attitudes,” or reactions to the quality of our own will toward others (FR 16). Strawson describes the reactive attitudes as the “attitudes...of involvement or participation in a human relationship” (FR 9). Without them, he argues,

---

<sup>1</sup> Throughout this paper, all in-text citations are to the 2008 reprint of “Freedom and Resentment” (“FR”).

there would “no longer [be] any such things as inter-personal relationships as we normally understand them” (FR 12), there would no longer be “anything that *we* could find intelligible as a system of human relationships, as human society” (FR 26). And, says Strawson, “it is just these attitudes themselves which fill the gap in the optimist’s account” (FR 25).

Strawson concludes “Freedom and Resentment” with the following remarks:

If we sufficiently, that is *radically*, modify the view of the optimist, his view is the right one. It is far from wrong to emphasize the efficacy of all those practices which express or manifest our moral attitudes, in regulating behaviour in ways considered desirable; or to add that when certain of our beliefs about the efficacy of some of these practices turn out to be false, then we may have good reason for dropping or modifying those practices. What *is* wrong is to forget that these practices, and their reception, the reactions to them, really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes... Indeed the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this. When we do remember this, and modify the optimist’s position accordingly, we simultaneously correct its conceptual deficiencies and ward off the dangers it seems to entail (FR 27).

The optimist’s error, then, is his failure to take into account all of the “facts as we know them” (FR 2). His view requires modification, not rejection. The optimist is right to appeal to the efficacy of our moral responsibility practices in regulating behavior in socially desirable ways, but wrong to focus only on that. He has failed to adequately describe our practices, to appreciate the “general framework of [reactive] attitudes” that forms “part of the general framework of human life” (FR 25). He has failed to grasp the role such attitudes play in explaining the efficacy of these practices at regulating behavior and in enabling

valuable interpersonal relationships. If the optimist were only to open his other eye to these facts, his view would become “the right one.” He would no longer be a one-eyed utilitarian.

But he would remain a utilitarian.

## II.

While Strawson’s reply to the pessimist is widely celebrated, his discussion of the optimist has attracted much less attention. Many interpret Strawson as a staunch opponent of utilitarianism, or, more generally, of consequentialist theories that attempt to justify our moral responsibility practices by reference to their beneficial consequences.<sup>2</sup> But a careful reading of Strawson’s discussion of the optimist suggests that Strawson was not opposed to the optimist’s consequentialist style of justification itself, but merely to the particular way he describes these practices and their consequences. It is true that Strawson defends these practices in part on grounds of their practical inescapability. We have no “choice in the matter” whether to hold one another responsible, since doing so flows from “our natural human commitment to ordinary inter-personal attitudes” (FR 14). But Strawson also insists that, if we could choose whether to maintain or abandon our moral responsibility practices, “then we could choose rationally only in the light of an assessment of the gains and losses to human life” (FR 14). Furthermore, though we cannot transform “the general structure” of these practices wholesale, we can make various “modifications internal to it” (FR 25). And whether or not we should make such changes depends, again, on “gains and losses to human life”: “when certain of our beliefs about the efficacy of some of these practices turn out to be false, then we may have good reason for dropping or modifying those practices” (FR 27).

No doubt, this interpretation of Strawson remains controversial. But from this point

---

<sup>2</sup> See, for example, Darwall (2006), Watson (1987), and Wallace (1994).

forward, my concern will be not so much to defend it as the best interpretation of Strawson's own position, as to take seriously Strawson's claim that the optimist's view, suitably modified to reflect an appreciation of the reactive attitudes, is "the right one," and to develop this revised optimistic position.<sup>3</sup> Strawson, I have suggested, never rejects the optimist's consequentialism, but rather sees his view as incomplete, and tries to give him "something more to say" (FR 4). In this paper, I say that something more. I adopt the perspective of Strawson's optimist, and show how an appreciation of the reactive attitudes can fill the gaps in the optimist's consequentialist defense of our moral responsibility practices, just as Strawson claims. In particular, I argue that an appreciation of the reactive attitudes should lead the optimist, first, to back away from the claim that we do or should hold people responsible in a forward-looking way—with the beneficial consequences of doing so in mind—in favor of the view that our moral responsibility practices *as a whole* are justified by their beneficial consequences (section III). It should lead the optimist, second, to recognize that such practices are efficacious at regulating behavior precisely because of their backward-looking character and the fact that we care about the attitudes others take towards us (section IV). And it should lead the optimist, third, to acknowledge that our moral responsibility practices are valuable not only because they effectively regulate behavior, but also because they enable interpersonal relationships (section V).

Filled in these three ways, optimism provides an attractive theory of moral responsibility—one that not only a utilitarian, but anyone who acknowledges the value of behavioral regulation and of interpersonal relationships can accept. And it moreover provides us with a fruitful framework for critically evaluating particular features of our moral

---

<sup>3</sup> Though see Miller (2014) and McGeer (2014) for more thorough attempts to defend a consequentialist interpretation of Strawson himself.

responsibility practices as well as proposed reforms: the two examples I consider are the role of moral luck in our current practices and the proposal that we should attempt to eliminate or suppress resentment and indignation (section VI). Yet it still faces a challenge from a certain sort of internal skeptic who claims that, the value of our moral responsibility practices notwithstanding, the standards internal to these practices commit us to thinking that nobody is ever morally responsible for anything. I return to this worry at the end of the paper, where I outline a two-pronged approach that optimists may appeal to in warding off this pessimistic rejoinder (section VII).

### III.

In the contemporary literature on moral responsibility, consequentialism tends to be represented by the specter of Strawson's one-eyed utilitarian. Worse, it tends to be represented by a particularly myopic variant of the view, as developed by J. J. C. Smart and his predecessor Moritz Schlick.<sup>4</sup> This is unfortunate. The myopic consequentialism of Smart and Schlick is highly implausible, but not because of its consequentialism. According to these views, whether or not a person is morally responsible for an action depends on the effect of holding her responsible for it. Someone is blameworthy for an action when blaming her for it has good consequences, praiseworthy for an action when praising her for it has good consequences. Blame and praise are thus rendered, on these accounts, "devices we calculatingly employ for regulative purposes" (FR 27).

On Smart's view, the primary benefit of holding others responsible lies in its regulatory function: "[t]o praise a class of actions is to encourage people to do actions of that

---

<sup>4</sup> Smart (1961); Schlick (1939, ch. 7).

class,” and to blame (or “dispraise”) them is to discourage this.<sup>5</sup> But Smart is a revisionist. His claim is not that we already hold one another morally responsible in this way, but that we should hold one another responsible in this way, since doing so would have good consequences.<sup>6</sup> Schlick’s view is similar, but differs in its emphasis and intent. He focuses not on the deterring and incentivizing function of holding people morally responsible, but on its educating and reforming function: on its ability to teach others moral rules and to influence their motives.<sup>7</sup> Indeed, for Schlick, blame is a form of punishment, and “[p]unishment is an educative measure, and as such is a means to the formation of motives.”<sup>8</sup> Furthermore, unlike Smart, Schlick does not take himself to be providing a revisionary theory of moral responsibility, but to be describing how we already use the concept of “responsibility.” On Smart’s normative theory, then, we ought to hold others morally responsible for their actions on the basis of forward-looking considerations: on the basis of what holding them morally responsible for their actions will (or is likely to) do. On Schlick’s descriptive theory, we already do this.

---

<sup>5</sup> Smart (1961, 305).

<sup>6</sup> Smart (1961, 304) writes that the “clear headed man” will hold people responsible in the way he suggests, but that “most men do not... praise and blame people in this dispassionate and clear-headed way” (305).

<sup>7</sup> Schlick (1939, 152). Compare William K. Frankena (1973, 74). Others who have emphasized the effect of our moral responsibility practices in developing people’s motives include Brandt (1969), Dewey (1922, ch. 4.4), and Vargas (2013, ch. 6). Those who, like Smart, focus more on deterrence and incentivization include Nowell-Smith (1948), Dennett (1984, ch. 7), and Sidgwick (1981, ch. 5).

<sup>8</sup> Schlick (1939, 152).

Let us begin with the inadequacy of the Schlickian descriptive theory. The theory faces three major problems, which I discuss here only briefly since they are already well known. First, it simply flies in the face of our phenomenology to think that we hold people responsible in a forward-looking way; as Victoria McGeer puts it, “when we praise and blame people, we’re not engaging in a kind of behavioural therapy, thinking of how our reactions might prod them into doing the things we approve of and avoiding the things that we don’t.”<sup>9</sup> This criticism should be intuitively persuasive to anyone who has ever held anyone responsible for anything. We certainly don’t seem to praise and blame in a forward-looking way. Yet one might object that our phenomenology misleads us. This brings us to the second major problem for Schlick’s view: that an overwhelming amount of experimental research suggests that we do blame and punish people on the basis of backward- rather than forward-looking considerations. In a recent paper, Fiery Cushman reviews some of this research, and provides a helpful summary:

Several lines of psychological research suggest a basic process of assigning blame and punishment... and in its details it is fundamentally [backward-looking]. When a harm occurs, we begin by seeking out individuals who are causally responsible. We then assess the harm-doers’ mental states at the time of their actions, determining whether they had a culpable mental state such as intent to harm or foresight. Finally, we assign punishment to the causally responsible parties in proportion both to the degree of the harm and the degree of their culpable mental state.<sup>10</sup>

---

<sup>9</sup> McGeer (2014, 70). Compare Bennett (2008), Wallace (1994), and Watson (1987).

<sup>10</sup> Cushman (2013, 346-347). See especially Darley et. al. (2000), Carlsmith et. al. (2002), and Carlsmith (2006), which suggest that modifying the expected effects of blame or punishment



At least in this case, then, our best evidence suggests that our phenomenology does not lie. We hold people responsible on the basis of what they have done and their quality of will in doing so, not on the basis of what the likely effects that holding them responsible for their actions will be. Schlick's theory of moral responsibility is empirically disconfirmed.

In addition to these phenomenological and empirical worries, the Shlickian view also faces conceptual difficulties. For example, Richard Brandt points out that the theory wrongly entails that "a person [is not morally responsible] if it would be a bad thing... to condemn him for it":

Suppose a vindictive and tyrannical king does something we think deserving of the most severe disapproval, but would be provoked to even more objectionable behavior if he were personally reproached in public, or even if the news came to him that he had been criticized in private. According to the theory, he is morally excused.<sup>11</sup>

---

tends not to impact blame or punishment behavior, as well as Fehr and Gächter (2002) and Nadelhoffer et. al. (2013), which suggest that people blame and punish even when doing so has no positive consequences. Note that while some of these studies focus primarily on punishment, the same basic story appears to hold for blame. For a useful review of some of the earlier literature emphasizing blame in particular, see Darley and Shultz (1990).

<sup>11</sup> Brandt (1969, 344). Brandt raises this objection against Smart, not Schlick. He furthermore complains that, on Smart's theory, "a person is properly held morally responsible for an action if he did not even perform it, provided that for some reason it is useful to perform this blaming act" (Ibid.) But as Arneson (2013) points out, the latter is not really an implication of Smart's (or Schlick's) theory. Even myopic consequentialists admit that a

But this is obviously wrong. Our attributions of moral responsibility do not vary in this way with the expected effect of holding others morally responsible; we do not excuse others for bad behavior because we think that, otherwise, they will act even worse. Schlick's descriptive theory therefore fails to adequately capture our concept of responsibility. As T.M. Scanlon puts it: "[t]he usefulness of administering praise or blame depends on too many factors other than the nature of the act in question for there ever to be a good fit between the idea of influenceability and the idea of responsibility which we now employ."<sup>12</sup>

Schlick's descriptive version of myopic consequentialism must therefore be rejected. And so we arrive at the first place where the reactive attitudes can "fill the gap in the optimist's account" (FR 25). A Strawsonian view on which we hold each other morally responsible in a backward-looking way, by taking up reactive attitudes towards others on the basis of what they have done and the quality of will their actions express, fits much better with our phenomenology and with the empirical research on blame and praise. It moreover avoids the conceptual problems that beset the Schlickian theory, not only cohering with our intuitive judgments about moral responsibility, but helping us to explain them.

By way of elaboration, consider, for example, the Strawsonian explanation of our practices of giving and taking exemptions and excuses. As we have seen, on Strawson's view, the reactive attitudes "rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves," where "[w]hat will, in particular cases, *count* as manifestations of good or 

---

person can only be morally responsible for actions she has performed. The view is forward-looking not because it denies this, but because it holds that a person's responsibility for the actions she performs depends on the consequences of holding her responsible for them.

<sup>12</sup> Scanlon (1988, 160).

ill will or disregard will vary in accordance with the particular relationship in which we stand to another human being” (FR 15). This allows us to understand exempting conditions as conditions that preclude a person from standing in the sort of relationships that admit the reactive attitudes: a person is exempted when they are “*incapacitated* in some or all respects for ordinary inter-personal relationships,” and this is why we exempt, for instance, those with severe cognitive impairments and young children (FR 13). As for excuses: we excuse people when they stand in the sort of relationships to us that admit the reactive attitudes, and when they do wrong, but without manifesting a poor quality of will. For example, though in many cases I will blame you for injuring me, if you do so accidentally, or because someone else compelled you to, or because you did not know that your action would harm me, I will not. Why? Because even though your action injured me, it manifested no ill will. So, on Strawson’s account, I have nothing to blame you for.

At least when it comes to the descriptive adequacy of our theory of moral responsibility, then, a Strawsonian approach clearly beats out a Schlickian view on which we hold people morally responsible in order to affect their future behavior. The optimist’s first modification is to acknowledge this, to accept a Strawsonian picture of our actual practices, and to clarify that she is not attempting to give a forward-looking account of how we hold people morally responsible in particular cases, but rather a forward-looking justification of our backward-looking moral responsibility practices taken as a whole.<sup>13</sup> The optimist, then, is a sort of moral responsibility positivist who believes that someone is morally responsible not when they meet some external, practice-independent standard, but when they meet certain standards internal to our moral responsibility practices, in much the same way that legal positivists believe that someone is legally responsible not when they meet some external,

---

<sup>13</sup> Compare Vargas (2013, ch. 6).

practice-independent standard, but when they meet certain standards internal to our existing legal practices.<sup>14</sup> Yet, at the same time, the optimist is not a relativist who thinks any moral responsibility practice is as good as any other. Her optimism consists in her conviction that our current moral responsibility practices are justified by their beneficial consequences.

#### IV.

This leads us back to Smart's normative theory, on which we should hold people responsible in a forward-looking way even though we currently do not. In response to the phenomenological and empirical arguments, Smart could reply that most of us fail to hold others responsible in a "dispassionate and clear-headed way."<sup>15</sup> And in response to the conceptual argument, Smart could reply that regardless of whether we wish to call his a theory of moral responsibility, it is certainly a theory of how we should respond to one another. From the optimist's perspective, the question of whether we should adopt Smart's proposal therefore turns not on our current standards of moral responsibility, but on what the consequences of adopting his revisionary proposal would be. It depends on whether a practice in which we held one another responsible on the basis of forward-looking considerations would have better results than our current practice of holding them responsible in reaction to the quality of their wills.

In making this comparison, the optimist may point to two important sorts of reasons to believe that our current moral responsibility practices have far better consequences than Smart's forward-looking alternative, and, indeed, that their great value lies precisely in their backward-looking character. The first pertains to the sort of instrumental considerations that

---

<sup>14</sup> Hart (1961).

<sup>15</sup> Smart (1969, 305).

utilitarians traditionally emphasize: the way that our current practices serve to regulate behavior in socially beneficial ways. Here, it is helpful to consider an analogy with punishment provided by John Rawls, who (playing the utilitarian) writes:

as one drops off the defining [backward-looking] features of punishment one ends up with an institution whose utilitarian justification is highly doubtful. One reason for this is that punishment works like a kind of price system: by altering the prices one has to pay for the performance of actions it supplies a motive for avoiding some actions and doing others. The defining features are essential if punishment is to work in this way; so that an institution which lacks these features... is likely to have about as much point as a price system (if one may call it that) where the prices of things change at random from day to day and one learns the price of something after one has agreed to buy it.<sup>16</sup>

Rawls's point here is that in order for punishment to effectively deter undesirable actions, people have to be able to anticipate the penalties that they will face if they perform those actions. But if a penal system instead apportioned punishment on the basis of forward-looking considerations, then punishment would depend on various factors pertaining to the effects of punishment at the time of punishment, and people would be unable to anticipate the penalty that would attach to their actions at the time of their performance. Indeed, attributing the view to Bentham, Rawls contends that:

if utilitarian considerations are followed penalties will be proportional to offenses in this sense: the order of offenses according to seriousness can be paired off with the order of penalties according to severity... This follows from the assumption that

---

<sup>16</sup> Rawls (1955, 12).

people are rational (i.e., that they are able to take into account the ‘prices’ the state puts on actions), [and] the utilitarian rule that a penal system should provide a motive for preferring the less serious offense.<sup>17</sup>

In other words, Rawls argues that consequentialist considerations justify a *backward-looking* system of punishment, and, in particular, one that will “insure that punishment is proportional to the offense.”<sup>18</sup> And the same argument can be used to justify a backward-looking system of moral responsibility. A Smartian forward-looking practice of moral responsibility would fail to effectively deter socially undesirable behavior for much the same reason as a forward-looking system of legal punishment. If people responded to others on the basis of what they believed the effects of those responses would be at the time of punishment, then individuals would often be unable to anticipate how others would respond to them, and they would furthermore have an incentive to “game the system” by putting themselves in situations where punishment or blame cannot have beneficial effects: to behave like Brandt’s tyrannical king. In short, the incentive structure of a forward-looking system of moral responsibility would be far less useful than the incentive structure provided by a backward-looking system. It would be worse on forward-looking grounds themselves.

This argument may be buttressed by some recent research on the subject in experimental economics and evolutionary theory. This work focuses on the role—the very large role—of punishment in securing and maintaining cooperation. In a recent paper, Shaun Nichols puts this research to work in defense of the reactive attitudes, focusing in particular

---

<sup>17</sup> Ibid., 12-13.

<sup>18</sup> Ibid., 12.

on the groundbreaking work of Ernst Fehr and his colleagues.<sup>19</sup> In a number of experiments, Fehr has explored the behavior of subjects playing public good games in laboratory settings. In these games, each player is provided with some set amount of money, and must decide how much of it to invest in a common pool. Subjects get to keep the money that they do not invest, while the amount that they collectively invest is multiplied by some factor, and then split evenly among the players. If all players invest all of their money, this results in the largest sum of money split evenly between the members of the group. But each player has an incentive to “free ride”: to keep her initial endowment for herself, while benefiting from the investments of others. The problem is immediate. Any individual can obtain more money by free riding than by contributing, but if all individuals free ride, everyone gets less than had they all contributed. But what is remarkable about what Fehr and his colleagues have shown is that if individuals in such games are granted the ability to punish free-riders—to pay some cost to reduce free-riders’ monetary payoffs—this allows them to solve the free-rider problem after all.

More specifically, in a series of experiments, Fehr and his colleagues have demonstrated the following (here I closely follow Nichols’ presentation):

- (i) Both participants in public goods games and third parties are often willing to punish free riders, even when doing so cannot yield the punisher any benefit. This tendency to punish is plausibly driven by the punishers’ emotional responses, and in particular by moral anger (resentment or indignation).
- (ii) When playing repeated games without punishment, cooperation tends to decay over time. When punishment is introduced, this drives cooperation “near

---

<sup>19</sup> Nichols (2007). The relevant studies are Fehr and Gächter (2000), Fehr and Gächter (2002), Fehr and Fischbacher (2004a), and Fehr and Fischbacher (2004b).

ceiling.”<sup>20</sup> For instance, in one treatment group, players were not able to punish until round 11 of the game. By round 10, contribution rates had dropped below 20%. In round 11, it climbed to 60%, and by round 14 to 90%. In many groups, contribution reached 100% by the final round.<sup>21</sup>

- (iii) Cooperation increases when individuals know that punishment is an option, even before anyone is punished. Fehr and Gächter suggest that this is because players anticipate that others will punish them if they defect. Indeed, players tend to be very accurate at predicting under what conditions others will punish them.

These results are interesting in their own right, but more to the point, they provide strong evidence that punishing others on the basis of backward-looking considerations does indeed play an important role in promoting and maintaining cooperation. And they furthermore reinforce the claim that in order for punishment to have this effect, people must be able to anticipate what behaviors will result in punishment—something that they could not do if others responded to them on the basis of forward-looking considerations. Still further support for this claim is provided by research conducted by Robert Boyd, Peter J. Richerson, and colleagues in the field of evolutionary theory that suggests roughly the same thing: that punishment, and specifically “targeted punishment” aimed at those who have defected, plays an integral role in stabilizing cooperation in large groups, since even groups whose members would be willing to cooperate in the absence of punishment are liable to “invasion” by free-riders if there are not at least a substantial number of targeted punishers in the population.<sup>22</sup>

---

<sup>20</sup> Nichols (2007), 419.

<sup>21</sup> This data is from Fehr and Gächter (2000).

<sup>22</sup> See Boyd et al. (2003), Boyd and Richerson (2005, chs. 9-10), Richerson and Boyd (2005, 199-201). Gaus (2011, ch. III., sect. 7) provides a helpful overview.



All in all, the evidence suggests that punishment is necessary to solve free-riding problems and to discourage socially detrimental behavior, not only in the experimental economics lab, but in real human communities as well.<sup>23</sup>

At this point one might protest that punishment and blame are very different things. Fair enough. But my argument does not rely on any claimed identity between punishment and blame. It relies only on the weaker claim that holding others responsible functions in one respect exactly like punishing them: that just as people do not like to be punished, and so are deterred by the prospect of punishment, people do not like being blamed, and so are deterred by the prospect of blame. So long as people prefer not to be blamed, and people expect to be blamed for their socially undesirable actions or attitudes, blame will therefore play precisely the same role as punishment in deterring defection and promoting cooperation. Indeed, Richerson and Boyd explain that in order for “punishment” to stabilize cooperation, it need not take a particularly active form: it may simply involve “reduced status” or “fewer friends.”<sup>24</sup> And as Strawson emphasizes, it is a “central commonplace” that people care deeply about what attitudes and intentions others hold towards them (FR 5). We do not like to be the target of the negative reactive attitudes, and are motivated to avoid being such targets, in much the way we are motivated to avoid active forms of punishment.

On top of all of this, there is a wide range of empirical evidence suggesting that backward-looking reactive attitudes like resentment and indignation play a key role in

---

<sup>23</sup> Though see Barrett (forthcoming) for a discussion of some of the limits of this sort of decentralized punishment.

<sup>24</sup> Richerson and Boyd (2005, 200)

motivating more active forms of punishment as well.<sup>25</sup> These sources of motivation play an especially important role generating credible threats of punishment in cases where perpetrators know that such punishment will have no beneficial effects after the wrong has occurred. We have already seen this in the case of Brandt's king, but to take a more mundane example of Robert H. Frank's: if it will cost you more to take me to court for illegally damaging your property than you stand to gain from winning the case, and you are motivated only by forward-looking considerations, then I face no credible threat of legal action and so will not be deterred from damaging your property. But if you will be motivated to take me to court out of backward-looking resentment for my crime, and I know this, then I do face a credible threat, and so will be deterred from damaging the property accordingly.<sup>26</sup>

Thus, we find the second place where the optimist must reach for the reactive attitudes to fill a gap in her story. The optimist must acknowledge that our disposition to form reactive attitudes, and more generally to care about the attitudes of others, is precisely what explains the efficacy of our moral responsibility practices in regulating behavior. She must reject Smart's revisionary proposal that we respond to one another in a forward-looking way on the grounds that this mode of interaction would be far less effective at regulating behavior than our current practices of holding each other morally responsible in a backward-looking way, and that it would therefore have worse consequences.

## V.

So the optimist may reject Schlick's descriptive theory while remaining a consequentialist,

---

<sup>25</sup> See Nichols (2007) and Haidt (2003) provide overviews. But see especially Izard (1977, ch. 13), Shaver et al. (1987), Pillutla and Murnighan (1996), and Hopfensitz and Reuben (2009).

<sup>26</sup> Frank (1988, 48).

and should reject Smart's normative theory on consequentialist grounds themselves. But there remains a third and final gap in her story, corresponding to the previously advertised second class of reasons why our moral responsibility practices are of great value. To repeat Strawson's earlier claim: our reactive attitudes and current practices of holding one another morally responsible are inextricably entwined with our participation in interpersonal relationships such that, without them, there would "no longer [be] any such things as interpersonal relationships as we normally understand them" (FR 12). There would no longer be "anything that *we* could find intelligible as a system of human relationships, as human society" (FR 26).

Of course, Strawson does acknowledge that we sometimes adopt the "objective attitude" toward others, where:

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 certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided (FR 9).

But to adopt this attitude toward another person is exactly to fail to treat her as a person with whom we stand in an ordinary adult interpersonal relationship. Insofar as Smart's myopic consequentialism requires us to abandon the backward-looking reactive attitudes and instead tailor our responses to others on the basis of the expected effects of such responses, it therefore requires us to take up the objective attitude toward others, and so is incompatible with the maintenance of our interpersonal relationships and with the complex structure of such relationships that forms an integral part of our society. To abandon our current moral responsibility practices in favor of Smart's revisionary proposal, then, would

be to turn our backs on society as we know it, and on our interpersonal relationships as we know them. But these relationships are highly valuable; it would be a grave loss to abandon them. And so it would be a grave loss to abandon our moral responsibility practices.

Perhaps no one is as eloquent as Susan Wolf in painting a bleak picture of a society “in which we all regarded each other with the objective attitude”:

We would applaud and criticize, say ‘thank you’ and ‘for shame’ according to whether our neighbors behavior was or was not to our liking. But these actions and words would have a different, shallower meaning.... Our praises would not be expressions of admiration and esteem; our criticisms would not be expressions of indignation or resentment. Rather, they would be bits of positive and negative reinforcement meted out in the hopes of altering the character of others.<sup>27</sup>

But that is not even the worst of it. As Wolf explains, “[t]he most gruesome difference between this world and ours would be reflected in our closest human relationships”:

We would still be able to form some sorts of association that could be described as relationships of friendship and love. One person could find another amusing or useful.... Nevertheless, I hope it is obvious that the words ‘friendship’ and ‘love’ applied to relationships in which admiration, respect, and gratitude have no part, might be said to take on a hollow ring. A world in which human relationships are restricted to those that can be formed and supported in the absence of the reactive attitudes is a world of human isolation so cold and dreary that any but the most

---

<sup>27</sup> Wolf (1981, 390-391).

cynical must shudder at the idea of it.<sup>28</sup>

The optimist will join Wolf in shuddering at this picture, and so supplement her initial account by pointing to *two* important sources of value that our moral responsibility practices provide: their regulation of behavior in socially desirable ways, and their role in enabling interpersonal relationships. She will thus strengthen her position by arguing that though particular instances of holding one another responsible may have bad consequences, and though we rarely attend to these consequences when engaging in blaming or praising behavior, our moral responsibility practices as a whole are justified by their consequences.

Now, Smart himself might find this justification unpersuasive, given his “extreme” or *direct* version of utilitarianism on which we must always evaluate individual actions by their consequences.<sup>29</sup> But our optimist instead accepts an *indirect* consequentialist theory on which we justify not particular actions, but existing practices, by their consequences.<sup>30</sup> She insists that focusing on the forward-looking consequences of particular instances of holding others responsible is too myopic, since a commitment to such a style of justification precludes us from obtaining the gains of a backward-looking system of moral responsibility. To provide a satisfying consequentialist defense of our moral responsibility practices she therefore zooms out, and focuses on the beneficial consequences of these practices as a whole. She recognizes that it is precisely by conforming to a practice of holding people responsible in a backward-looking way that we are able to do so much good going forward.

---

<sup>28</sup> Wolf (1981, 391). Compare, for example, Bennett (2008), McGeer (2014), and especially Shabo (2012).

<sup>29</sup> Smart (1956)

<sup>30</sup> See especially Rawls (1955), Miller (2009), and Susan Wolf (2016).

## VI.

To this point, we have seen how the position of the optimist can be modified to accommodate the reactive attitudes, and how optimism can therefore retain its distinctive consequentialist commitment while casting off the myopia of one-eyed utilitarianism. In particular, we have seen that the optimist can adopt a Strawsonian understanding of our actual moral responsibility practices, and can tell a powerful story about why these practices are justified—a story, that, as I have mentioned, anyone who sees the value of regulating behavior and in enabling interpersonal relationships may accept. Yet the optimist defends the status quo, and this might seem to open her to the charge of being overly complacent or conservative. Surely we want our theory of moral responsibility to provide us with a perspective from which we can not only describe and justify, but also criticize and propose reforms to our existing practices. In fact, however, this is precisely the perspective that the optimist provides. Her optimism extends far enough for her to reject revolutionary views on which we should do away with our reactive attitudes altogether, but, as Strawson notes, she does not deny the desirability of making “modifications internal to” our moral responsibility practices (FR 25). Indeed, perhaps the greatest appeal of optimism about moral responsibility is its ability to provide us with precisely the critical perspective we need.

In approaching issues of moral responsibility, the optimist has available to her two perspectives. The first is a perspective internal to our moral responsibility practices; the second is one external to them.<sup>31</sup> When the optimist takes the first perspective, she is

---

<sup>31</sup> This distinction should not be confused with Strawson’s distinction between the participant attitude and the objective attitude, since, from the optimist’s external standpoint, she still sees people as embedded in interpersonal relationships. It is closer to Hart’s (1961) distinction between the external and the internal points of view of a legal system, or Rawls’s

concerned with examining the contours of our current practices, with determining when it is appropriate to hold others responsible (or perhaps to make attributions of responsibility) according to standards internal to these practices. In examining, for example, the issue of moral luck, she will consider various thought experiments and real cases, in order to examine her reactions to these cases.<sup>32</sup> She will consider cases where two agents perform the same action, with the same quality of will, yet these actions have different consequences due to factors beyond the agents' control—such as a case in which one reckless driver unluckily hits and kills a child, and another equally reckless driver luckily avoids harming anyone. And she will explore how her reactions to these cases fit with her reactions to other cases, attempting to determine the features of the cases in virtue of which she has the reactions that she does. She will, in other words, conduct herself very much like the majority of moral responsibility theorists treating these issues, but without thinking that her reactions to these cases track anything other than the standards internal to our current practice. She will treat her own reactions as evidence for what is appropriate or inappropriate within this practice—since she is herself a competent participant in these practices—and compare them to the reactions of other moral responsibility theorists, and to experimental evidence about how individuals do in fact form such judgments, in order to come to a more adequate description of our existing practices and the standards internal to them.

Yet the optimist may also fall back on an external perspective, one concerned not

---

(1955, 3) distinction between the perspective of “justifying a practice” and that of “justifying a particular action falling under it.” Compare also Hare’s (1981) distinction between the critical and intuitive levels of moral thinking.

<sup>32</sup> On moral luck, see Williams (1981) and Nagel (1979). Here I focus on “resultant” moral luck in particular.

with the internal appropriateness conditions of our practices, but instead with these practices' consequences, and, in particular, with their ability to regulate behavior and enable valuable relationships. Rather than thinking that there is some practice-independent fact of the matter whether, say, people's degree of responsibility can vary with instances of moral luck (such that the relevant question is whether our practices accurately track this fact), our optimist will instead be concerned with identifying the benefits and costs of our practice of holding people responsible partially on the basis of moral luck, and will examine whether there are feasible ways of reforming this feature of our practice to make it have better consequences. For example, our optimist might begin with the thought that if the goal of our practices is to deter bad behavior, then it makes little sense to attach more severe blame to behavior that has worse consequences due to factors beyond the agent's control, since we cannot deter agents from bringing about outcomes that are outside their control. But she might then note that our best evidence about people's quality of wills comes from their actions and those actions' outcomes, and so conclude, with Victor Kumar, that our epistemic limitations preclude us from reliably "assigning responsibility purely on the basis of opaque intentions," such that "[a]ssigning responsibility partly on the basis of outcomes is more reliable...and thus is able to regulate behavior more effectively."<sup>33</sup> Following Kumar, she might also point to psychological studies suggesting that people are "better capable of *moral learning* when punishment is matched to outcomes rather than intentions"—that people are better at learning moral rules and modifying their behavior accordingly in a system that rewards positive outcomes and punishes negative ones rather than one that focuses exclusively on mental states.<sup>34</sup> Or, turning from behavioral regulation to interpersonal

---

<sup>33</sup> Kumar (2019, 998).

<sup>34</sup> Kumar (2019, 1004). See also Cushman (2013) and Martin and Cushman (2016).



relationships, she might note that it would put undue strain on our interpersonal relationships if we were required to respond in the same way, say, to people who have recklessly run over our children as to those who have equally recklessly just missed them, given the much greater “significance” of the former event on our relationship with the driver.<sup>35</sup> And so she might conclude that, on balance, our practice of attributing moral responsibility partially on the basis of moral luck is justified by its effect on behavioral regulation and on interpersonal relationships, her initial reservations notwithstanding.

To be clear, my purpose here is not to come down one way or the other on the issue of moral luck, but rather to illustrate how a suitably revised optimistic position allows us to critically assess our moral responsibility practices without losing sight of their rich texture. In some cases, optimism may allow us to justify features of our practice that seem otherwise unjustified on grounds of their arbitrariness, say, the precise threshold at which we deem that someone moves from innocently unobservant to culpably negligent: as with our legal practices, our moral responsibility practices need “efficiently determinable thresholds,” and it sometimes matters more that we draw a sharp line somewhere (at least within some satisfactory range) than that we draw it in any particular point.<sup>36</sup> Yet in other cases, optimism is more uncompromising than other theories of moral responsibility, always demanding an external justification for features of our practices rather than accepting them merely because they mesh with our intuitions. Optimism does not imply complacency. Though the optimist believes that our moral responsibility practices are justified as a whole, she is highly critical of particular features of them, always on the lookout for feasible ways of reforming them.

Consider, for example, the claim made by critics of our moral responsibility practices

---

<sup>35</sup> Scanlon (2008, 138).

<sup>36</sup> Dennett (1984, 162).

such as Derk Pereboom that even though we should not give up the reactive attitudes altogether, we would be better off without reactive attitudes like resentment or indignation, or at least with a whole lot less of them. The basic style of argument here is to argue, first, that these attitudes bring significant costs, and, second, that we could achieve their benefits through alternative attitudes that play similar functional roles. In particular, Pereboom and others point out that resentment and indignation may be used to enforce bad norms or laws, may motivate excessively harsh punishment of good norms and laws, and may damage interpersonal relationships.<sup>37</sup> They then argue that other attitudes such as “moral sadness and sorrow—accompanied by a resolve for fairness and justice, or to improving personal relationships—[can] serve societal and personal relationships as well as resentment and indignation,” but without the associated costs.<sup>38</sup>

In evaluating such proposals for reform, the optimist’s first move is to distinguish our moral responsibility practices from the social rules (norms and laws) that these practices serve to stabilize. She will recognize that there are countless examples of communities enforcing pernicious rules, holding people to bad expectations, and she will be concerned to determine whether a reform is more wisely aimed at changing these underlying rules and associated expectations than at changing the way we hold people responsible in general. The reactive attitudes are triggered by violations of expectations, and the problem is often that we are holding people to the wrong expectations rather than that we are holding people to expectations in the wrong way. Of course, these problems are interconnected. As Kate Manne has recently argued with respect to pernicious gender norms in particular, one reason these norms are so sticky is that their enforcement is often genuinely motivated by reactive

---

<sup>37</sup> Pereboom (2001, ch. 7), Pereboom (2014, ch. 6). Compare Caruso (2019).

<sup>38</sup> Pereboom (2014, 148).

attitudes, such that blaming those who depart from these expectations feels, from the inside, “righteous: like standing up for oneself or for morality...like a moral crusade, not a witch hunt...not in the spirit of hating women but, rather, of loving justice.”<sup>39</sup> But the point remains that just as one can call particular criminal laws into question without calling into question the structure of the judicial system, one can call particular norms and laws into question without calling into question the structure of our moral responsibility system. It is important that our moral responsibility practices are not so intransigent that they preclude moral progress. But in the absence of good evidence that our current system tends towards the enforcement of worse rules than some alternative system that replaces resentment and indignation with sadness and resolve, the fact that we sometimes attach resentment and indignation to bad norms is neither here nor there.

Still, the optimist should take seriously the claim that we would be better served by adopting alternative attitudes to resentment and indignation, not because such attitudes might be used to enforce bad norms, but because they may lead to excessive punishment of those who violate good norms or to the fracturing of relationships. For reasons we explored above, she will be hesitant to throw her support behind a system that does without the backward-looking attitudes of resentment and indignation altogether in favor of forward-looking attitudes like “resolve.” Though it might be nice if we could live in a world in which individuals would cooperate out of moral motivation and resolute encouragement from others, free-riders and other morally unsavory types will always pop up, and a resolve to affect positive change is not enough to stop them from “invading” (spreading throughout the community) and destabilizing cooperation. After all, as we saw when discussing Smart’s revisionary proposal, blame motivated in this forward-looking way is less apt to deter, not

---

<sup>39</sup> Manne (2018, 20).

only because it is less predictable, but also because it cannot as effectively motivate punishment of wrongdoing, or produce any credible threats of punishment at all in cases where perpetrators know that such punishment will have no beneficial effects after the wrong has occurred. Yet at the same time, our optimist should acknowledge that our current practices may very well involve too much resentment and indignation. Indeed, while the evolutionary models mentioned earlier suggest that a community that lacks punishers is prone to invasion by free-riders, they also allow that communities that are full of punishers are prone to engage in more punishment than is socially beneficial, rendering them prone to invasion from less punitive types.<sup>40</sup> Perhaps we could all benefit from this latter sort of invasion, or at least from some lessons in anger management.

Finally, when it comes to interpersonal relationships, the optimist must admit that the jury is still out. Though we are all aware of particular cases where too much resentment or indignation damages a relationship, the debate over whether valuable relationships can persist without any (disposition to) resentment or indignation has, to this point, remained largely speculative and *a priori*.<sup>41</sup> Yet the extent to which resentment and indignation are required in interpersonal relationships is an empirical question, and one that, as far as I am aware, has received little sustained attention. Here, Pereboom points us in the right direction when he suggests that we turn our attention to real world “communities in which training and teaching methods are employed to diminish resentment and indignation” in order to examine how successful such groups actually are at suppressing such emotions and the

---

<sup>40</sup> See again Boyd et al. (2003), Boyd and Richerson (2005, chs. 9-10), and Gaus (2013, ch. III, sect. 7).

<sup>41</sup> See, for example, the exchange between Shabo (2012) and Milan (2016).

consequences of this suppression.<sup>42</sup> This is the sort of careful empirical work that optimism requires us to engage in when evaluating the feasibility and desirability of potential changes to our moral responsibility practices.

## VII.

We have spent a while with the optimist, but I cannot end this paper without returning to the pessimist. As Strawson argues, optimists about moral responsibility need pay no mind to the sort of external skepticism about moral responsibility that points to practice-independent standards about what is required for freedom or responsibility and insists that we do not meet these requirements. Optimists deny that any such standards exist. But, as many since Strawson have responded, optimism does leave the door open to a certain sort of internal skeptic who revives this objection by claiming that the standards internal to our current moral responsibility practices commit us to requirements for moral responsibility that we cannot meet.<sup>43</sup> In other words, someone might claim that the very standards that are already embedded in our practices imply that it is only appropriate to hold others responsible if they have a form of libertarian free will that they cannot have in a deterministic world, or worse, in any world, and thus that no one is ever responsible for anything.<sup>44</sup>

In responding to this sort of skepticism, the optimist may appeal to a two-pronged approach. The first is to rebuff skeptics on their own terms. One of the benefits of

---

<sup>42</sup> Pereboom (2014, 149).

<sup>43</sup> See especially Russell (2013).

<sup>44</sup> One might think, for example, that our ordinary practices of excusing and exempting commit us to viewing determinism as excusing all actions or exempting all people. R. Jay Wallace (1996, 16) calls this strategy of arguing for skepticism the “generalization strategy.”

Strawsonian quality of will theories of moral responsibility is that they purport to explain the internal standards of our practices in a way that doesn't require anything so metaphysically fancy as libertarian free will. If we hold people responsible in reaction to the quality of their wills, then, *prime facie*, determinism poses no threat: it can't change the fact that people's actions do in fact reflect different qualities of will. But of course, the devil is in the details here, and to engage properly with this form of skepticism the optimist must provide a whole different argument than the one I have provided here, one launched entirely from the internal perspective. Thankfully, there is already a huge literature on this subject, and there is nothing preventing the optimist from tapping into this literature in developing the first prong of her response to the skeptic.<sup>45</sup>

Even if the internal skeptic ultimately carries the day, however, the optimist may fall back on a second prong. Pereboom and other theorists who believe we ought to abandon the reactive attitudes of resentment and indignation are typically skeptics about moral responsibility who believe that such attitudes rest on false beliefs or irrationality. Their conviction that we ought to abandon these attitudes is driven in the first instance by their skepticism, with their contention that doing so would lead to positive consequences arising only as an attempt to ward off the further worry that this abandonment would be highly detrimental. But from the optimist's perspective, the question of whether we ought to revise our moral responsibility practices depends entirely on the value of such revisions. In effect, then, the skeptic points to one source of disvalue so far missing from the optimist's account: the disvalue of living a life in which we have false beliefs or manifest certain forms of irrationality. Yet in most cases, it is hard to believe that this sort of disvalue could really outweigh the great value of social regulation or of interpersonal relationships. If we can

---

<sup>45</sup> McKenna and Pereboom (2016) provide a helpful starting place.

modify our practices to avoid a reliance on false beliefs or irrationality—perhaps even at some minor cost to their regulative and relationship-maintaining function—then the optimist should be all for this.<sup>46</sup> But if we can't, then the optimist must conclude that we would do best to ignore the false beliefs or irrationality that underlie our practices, and carry on as if we had never noticed.<sup>47</sup>

Admittedly, there would be something regrettable about this last outcome, just as there is always something regrettable about cases when value conflicts prove to be irreconcilable. It would turn out that we humans can live together on cooperative terms while participating in valuable interpersonal relationships only by living lives of self-deception or irrationality. As for myself, I tend to side with those compatibilists who believe the threat of internal skepticism can be warded off at the first prong, and so feel no deep regret at the human condition. But perhaps I'm just an optimist at heart.<sup>48</sup>

#### Works Cited

- Arneson, Richard. "The Smart Theory of Moral Responsibility and Desert." In *Desert and Justice*, ed. Serena Olsaretti (New York: Oxford University Press, 2013).
- Barrett, Jacob. "Punishment and Disagreement in the State of Nature." *Economics and Philosophy* (forthcoming): 1-21. <https://doi.org/10.1017/S0266267119000233>.
- Bennett, Jonathan. "Accountability (II)" in *Free Will and Reactive Attitudes*, ed. Michael McKenna and Paul Russell (Burlington: Ashgate, 2008).

---

<sup>46</sup> On such revisionism, see Vargas (2013, ch. 3)

<sup>47</sup> Compare Smilansky (2000).

<sup>48</sup> For helpful comments and discussion, I would like to thank Jerry Gaus, Michael McKenna, Sarah Raskoff, Hannah Tierney, Robert Wallace, and two anonymous referees.

Boyd, Robert and Peter J. Richerson. *The Origin and Evolution of Cultures*. (New York: Oxford University Press, 2005).

Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. "The Evolution of Altruistic Punishment." *Proceedings of the National Academy of Sciences of the United States of America* 100 (2003): 3531-3535.

Brandt, Richard B. "A Utilitarian Theory of Excuses." *The Philosophical Review* 78 (1969): 337-361.

Carlsmith, Kevin M. "The Roles of Retribution and Utility in Determining Punishment." *Journal of Experimental Social Psychology* 42 (2006): 437-451.

Carlsmith, Kevin M., John M. Darley, and Paul H. Robinson. "Why do we punish? Deterrence and Just Deserts as Motives for Punishment." *Journal of Personality and Social Psychology* 83 (2002): 284-299.

Caruso, Gregg D. "Free Will Skepticism and Its Implications: An Argument for Optimism." In E. Shaw, D. Pereboom, and G. D. Caruso (eds.), *Free Will Skepticism in Law and Society: Challenging Retributive Justice* (Cambridge: Cambridge University Press, 2019).

Cushman, Fiery. "The Role of Learning in Punishment, Prosociality, and Human Uniqueness." In K. Sterelny, R. Joyce, B. Calcott, and B. Fraser (eds.), *Cooperation and its Evolution* (Cambridge: MIT Press, 2013).

Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. (Cambridge: Harvard University Press, 2006).

Darley, John M, Kevin M. Carlsmith, and Paul H. Robinson. "Incapacitation and Just Deserts as Motives for Punishment." *Law and Human Behavior* 24 (2000): 659-683.

Darley, John M. and Thomas R. Shultz. "Moral Rules: Their Content and Acquisition." *Annual Review of Psychology* 41 (1990): 525-556.



Dennett, Daniel C. *Elbow Room: The Varieties of Free Will Worth Wanting* (Oxford: Clarendon Press, 1984).

Dewey, John. *Human Nature and Conduct*. (New York: Henry and Holt, 1922).

Fehr, Ernst and Urs Fischbacher. "Social Norms and Human Cooperation." *Trends in Cognitive Sciences* 8 (2004a): 187-190.

Fehr, Ernst and Urs Fischbacher. "Third Party Punishment and Social Norms." *Evolution and Human Behavior* 25 (2004b): 63-87.

Fehr, Ernst and Simon Gächter. "Altruistic Punishment in Humans." *Nature* 415 (2002): 137-140.

Fehr, Ernst and Simon Gächter. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90 (2000): 980-994.

Frank, Robert H. *Passions Within Reason* (New York: W. W. Norton and Co, 1988).

Frankena, William K. *Ethics* (Englewood Cliffs: Prentice Hall, 1973).

Gaus, Gerald. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World* (New York: Cambridge University Press, 2011).

Haidt, John. "The Moral Emotions." In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (eds.), *Handbook of Affective Sciences*. New York: Oxford University Press, 2003.

Hare, R. M. *Moral Thinking: Its Levels, Method, and Point* (New York: Oxford University Press, 1981).

Hart, H.L.A. *The Concept of Law* (New York: Oxford University Press, 1961).

Hopfensitz, Astrid and Ernesto Reuben. "The Importance of Emotions for the Effectiveness of Social Punishment." *The Economic Journal* 119(2009): 1534-1559.

Izard, Carrol E. *Human Emotions*. New York: Plenum Publishing Co.

Kumar, Victor. "Empirical Vindication of Moral Luck." *Noûs* 53 (2019): 987-1007.

Strawson, P.F. "Freedom and Resentment." In *Freedom and Resentment and Other Essays* (New York: Routledge, 2008).

Manne, Kate. *Down Girl: The Logic of Misogyny* (New York: Oxford University Press, 2018).

Martin, Justin W. and Fiery Cushman. "The Adaptive Logic of Moral Luck." In J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy* (Chichester: John Wiley and Sons, 2016).

McKenna, Michael and Derk Pereboom. *Free Will: A Contemporary Introduction*. (New York: Routledge, 2016).

McGeer, Victoria. "P.F. Strawson's Consequentialism." *Oxford Studies in Agency and Responsibility* 2 (2014): 64-92.

Milan, Per-Erik. "Reactive Attitudes and Interpersonal Relationships." *Canadian Journal of Philosophy* (46): 101-122.

Miller, Dale E. "'Freedom and Resentment' and Consequentialism: Why 'Strawson's Point' Is Not Strawson's Point." *Journal of Ethics & Social Philosophy* 8 (2014): 1-24.

Miller, Richard B. "Actual Rule Utilitarianism." *Journal of Philosophy* 106 (2009): 5-28.

Nadelhoffer, Thomas, Saeideh Heshmanti, Deanna Kaplin, and Shaun Nichols. "Folk Retributivism and the Communication Confound." *Economics and Philosophy* 29 (2013): 235-261.

Nagel, Thomas. "Moral Luck" in *Mortal Questions* (New York: Cambridge University Press, 1979).

Nichols, Shaun. "After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes." *Philosophical Perspectives* 27 (2007): 405-428.

Nowell-Smith, P. "Freewill and Moral Responsibility." *Mind* 57 (1948): 45-61.

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

- Pereboom, Derk. *Free Will, Agency, and Meaning in Life* (New York: Oxford University Press, 2014).
- Pillutla and Murnighan. "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers." *Organizational Behavior and Human Decision Processes* 68 (1996): 208-224.
- Rawls, John. "Two Concepts of Rules." *The Philosophical Review* 64 (1955): 3-32.
- Richerson, Peter J. and Robert Boyd. *Not By Genes Alone: How Culture Transformed Human Evolution* (Chicago: University of Chicago Press, 2005).
- Russell, Paul. "Responsibility, Naturalism, and 'The Morality System'." *Oxford Studies in Agency and Responsibility* 2 (2013): 184-204.
- Scanlon, T.M. "The Significance of Choice." In *The Tanner Lectures in Human Values*, vol 7., ed. Sterling M. McMurrin (Salt Lake City: University of Utah Press, 1988).
- Scanlon, T. M. *Moral Dimensions: Permissibility, Meaning Blame* (Cambridge: Harvard UP, 2008).
- Schlick, Moritz. *Problems of Ethics*, trans. David Rynin (New York: Prentice-Hall, 1939).
- Shabo, Seth. "Incompatibilism and Personal Relationships: Another Look at Strawson's Objective Attitude." *Australasian Journal of Philosophy* (90): 131-147.
- Shaver, Phillip, Judith Schwartz, Donald Kirson, and Cary O'Connor. "Emotion knowledge: Further exploration of a prototype approach." *Journal of Personality and Social Psychology* 52 (1987): 1061-1086.
- Sidgwick, Henry. *The Methods of Ethics*, 7<sup>th</sup> ed. (Indianapolis: Hackett, 1981).
- Smart, J.J.C. "Extreme and Restricted Utilitarianism." *The Philosophical Quarterly* 6 (1956): 344-354.
- Smart, J.J.C. "Free-Will, Praise, and Blame." *Mind* 70 (1961): 291-306
- Smilansky, Saul. *Free Will and Illusion* (New York: Oxford University Press, 2000).

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

Wallace, R. Jay. *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press, 1994).

Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In F. Schoeman (ed.), *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987)

Williams, Bernard. "Moral Luck." In *Moral Luck* (New York: Cambridge University Press, 1981).

Wolf, Susan. "The Importance of Free Will." *Mind* 90 (1981): 386-405.

Wolf, Susan. "Two Concepts of Rule Utilitarianism." *Oxford Studies in Normative Ethics* 6 (2016): 123-144.