The Basic Argument and Modest Moral Responsibility

1. Introduction

There is a traditional argument made explicit by Galen Strawson that moral responsibility is impossible.[1] The gist of this argument goes something like this:

1. To be morally responsible for her decision to *x*, an agent must be morally responsible for the state of mind that issues in *x*.
2. To be morally responsible for the state of mind that issues in *x*, an agent must be morally responsible for the decisions that caused her to have that state of mind.
3. To be morally responsible for *those* decisions, however, she would be to be morally responsible for the state of mind that issued in those decisions.
4. But if we keep tracing the etiology of the decision to *x* we'll fail to find a state of mind for which the agent is not morally responsible only if an infinite regress of mental states is possible or the agent is *causa sui*.
5. An infinite regress of mental states is impossible.
6. No (human) agent is *causa sui*.
7. So, the agent is not morally responsible for *x*.

And worrisomely, (7) will generalize to the further claim that no one is ever morally responsible for anything.

---

[1] Galen Strawson, "The Impossibility of Moral Responsibility," *Free Will 2nd Edition*, ed. Gary Watson, Oxford University Press (2003): 212-228.

This argument, it seems to me, is quite intuitive. At least, from my perspective, it's fairly clear why someone might feel the force of these premises and the inferences that link them. The conclusion, however, is not very intuitive. It seems, in the natural course of events, that we take ourselves to be morally responsible for our actions. In this way, the Galen Strawson-style argument is like many skeptical arguments in other domains: from intuitive premises and seemingly valid inferences, we reach a very startling and seemingly implausible conclusion. Unlike many skeptical arguments, however, I think Galen Strawson's argument can be given a reading on which it is very plausible that its conclusion is sound. Though as we will see, in order to develop a sound version of this argument, we'll have to give up on the idea that it has a genuinely revisionary conclusion, since we'll have less reason to care about the property that Strawson identifies with being morally responsible (i.e., the property that figures in (1) – (7)) than we might initially suppose. And this dulls the skeptical force of Strawson's argument considerably. This point, which is opaque now, will (I hope) become increasingly clear.

2. Ultimate Moral Responsibility

Premise (1) of the above argument tells us that a necessary condition on, say, my being morally responsible for my decision to leave work early after the boss has told me that I needed to stay late to finish a project, is that I am morally responsible for the state of mind that issues in this decision. More precisely, it says that I am morally responsible for this decision only if the beliefs, desires, cares, loves, values, intentions, or character traits that produced the decisions are mental states or dispositions for which I am morally responsible. (In this way, the notion of "state of mind" is to be understood capaciously.)

Now, at first blush, you might think this claim seems implausible. That is, you might want to insist that it is not a necessary condition on my being responsible for cutting out early that I'm responsible for the antecedent mental states that caused me to cut out early. As John Martin Fischer suggestively puts it, it's actually "metaphysical megalomania" that leads us to this conclusion—not a sound theory of moral responsibility.[2] And though I will ultimately want to agree with Fischer on this point, I also want to claim that it's not so maniacal (metaphysically or otherwise) to accept a suitably explicated version of (1).

To help see this, suppose, as Galen Strawson does, that *what it is* for an agent *S* to be morally responsible for a wrongful action *x* is for *S* to be such that *S* would deserve eternal torment in hell on the basis of *x*-ing.[3] Though this might seem extreme, the idea here captures something intuitive. After all, God's punishment, one might think, would necessarily be just. Consequently, it would seem that by specifying the kind of control that would warrant this form of punishment,[4] we have thereby specified the requirements of morally responsible agency in a way that abstracts away from the contingent infelicities in our own moral responsibility practices.

On this conception of *what it is* to be morally responsible, (1) seems fairly plausible. In response to Fischer's charge of "metaphysical megalomania," it therefore seems fair to point out that when the property of being morally responsible is itself so metaphysically and normatively loaded, the conditions on instantiating the property of being morally responsible are plausibly thought to be similarly loaded.

---

[2] John Martin Fischer, "Sourcehood: Playing the Cards that are Dealt You," *Deep Control*, Oxford University Press (2011): 163-187. Interestingly, Nietzsche suggests the same thing in *Beyond Good and Evil* I.21.
[3] Galen Strawson, (2003).
[4] Of course, we'd also have to specify the epistemic requirements on just divine punishment, but since just about everyone else (excepting George Sher, *Who Knew? Responsibility Without Awareness*, Oxford University Press (2009)) glosses over these requirements, I will as well.

Of course, you still might ask for more. In particular, you might want to know *why*, even on this picture of moral responsibility (which Galen Strawson calls *ultimate* or *true* moral responsibility) it is plausible to think that (1) is true. That is, even if you aren't generally skeptical about the metaphysical assumptions driving (1), you still might think we still need a reason that goes beyond "it's intuitive" or "it's plausible" to accept (1).

On one way of further motivating (1), we can begin by noting that given ultimate moral responsibility's constitutive connection to the notion of eternal torment in hell, the practical stakes are quite high on judgments of moral responsibility. Now notice that if I can be morally responsible for cutting out of work early even if that action issues from mental states and dispositions for which I am not morally responsible, then seemingly, I could be morally responsible for that action even if some demon or nefarious neuroscientist programmed me to have the very mental states and dispositions that caused my action. But suppose that some demon or (some really mean-spirited neuroscientist) *had* programmed me to have the mental states and dispositions that issued in me lazily cutting out of work early. Would I be morally responsible for that action? *Ultimately* morally responsible for that action? Deserving of hell?

With the threat of eternal torment hanging over my head, it's plausible, I think, to judge that I wouldn't be responsible. After all, could it really be fair for God to eternally damn me given the presence of the manipulator in the causal history of my action? I can't help but to think that I'm not alone in thinking that it couldn't be.

Of course, even if there's widespread agreement about this judgment, we'll still want to know why it wouldn't be fair for God to treat me in this way. A natural answer is simply

that in such a case, I'm not really the source of the action.[5] For although the action is my own in some sense (it's not as if it was one my co-workers who left early), it issued from states and dispositions that are not my own and for which I wasn't morally responsible. Given this, eternal torment seems not just unfair, but *manifestly so*. Therefore, if we follow Galen Strawson in thinking that the relevant notion of moral responsibility is *ultimate moral responsibility*, then it seems that we have good reason to accept (1): anything less would be to think that an agent could deserve eternal torment even if she wasn't the source of her action—even if she was programmed by God, demons, or those nefarious neuroscientists who (primarily) only live in philosophical thought experiments.

And if we accept (1) for this reason, then we should similarly accept (2). After all, sourcehood of the sort that might merit eternal torment would seem to also require that we are the source of the mental states and dispositions that issue in our actions. And once we have (1) and (2), the only further point of contention is premise (5), which states, in essence, that we are not self-caused prime movers. Although reasonable people can disagree about such things,[6] I think that neither (5) nor (6) is credible given the success of our best naturalistic theories of human action and motivation.[7] Consequently, we can accept the argument's conclusion—at least when it's made explicit that saying that no agent is ever morally responsible for their actions only commits us to the claim that *no agent is such that she instantiates a property in virtue of which she is deserving of being eternally tormented in hell.*

---

[5] Here I'm following Derk Pereboom and Patrick Todd, among others. See especially Pereboom, *Living Without Free Will*, (2001) and Todd, "Manipulation and Moral Standing: An Argument for Incompatibilism," *Philosophers' Imprint*, (2012).

[6] See especially Roderick Chisholm, "Human Freedom and the Self," *Free Will 2nd Edition*, ed. Gary Watson, Oxford University Press (2003): 26-37.

[7] To be clear: I am not defending (6) by claiming that it is literally incoherent or metaphysically impossible (in this way, I am parting ways with Strawson's own defense of the Basic Argument). Rather, I am offering an argument to the best explanation: viz., if our best scientific theories of the mind are incompatible with our status as prime movers, then we have good reason to doubt that status. For more in defense of this claim, see Derk Pereboom (2001).

3. Modest Moral Responsibility

Of course, even if you don't think that I'm *ultimately* morally responsible for leaving work early, without assuming anything about whether I am (or am not) morally responsible for the mental states and dispositions that led to my truancy, you might think that my co-workers' resentment, when directed at my decision to be lazy and cut out early, is appropriate. So too, you might think that the sharp warning I receive from my boss is also deserved. I did leave them all in a lurch with no excuse, after all.

But how can I be deserving of anything if I'm not ultimately morally responsible? Well, simply put, not only does it not follow from the fact that I am not deserving of eternal torment in hell that I am not deserving of any negative responses to my behavior, it's actually quite plausible that the conditions under which eternal torment are appropriate are importantly different than the conditions under which emotional responses like resentment or office discipline are appropriate. Indeed, we might expect this difference is built directly into the differences between the kind of control that would underwrite the propriety of eternal torment and the kind of control that would underwrite the propriety of reactive attitudes like resentment.[8] And because the former response involves higher practical stakes than does the latter, it stands to reason that the exact measure of control required for the former would not be identical to that required by the latter.

Moreover, resentment and related attitudes like gratitude (in positive case) or indignation are deeply—perhaps essentially—connected to our participation in ordinary interpersonal relationships, like our relationships with co-workers or bosses. This is because these attitudes reflect our concern for mutual respect, which is the normative ideal of these

[8] Here I'm invoking *Peter* Strawson's notion of the set of emotions that are specifically tied to quality of others' will. In addition to resentment, the set of reactive attitudes includes indignation, guilt, gratitude, esteem, pride, hurt feelings, (perhaps) love, etc. For more, see P. F. Strawson, "Freedom and Resentment," *Free Will 2nd Edition*, ed. Gary Watson, (2003): 72-93.

sorts of reciprocal relationships. Accordingly, it's plausible that an agent could be an apt target of these attitudes solely in virtue of those agential capacities that are necessary for participating in reciprocal interpersonal relationships more generally. To see this, consider the following argument:

8. Reciprocal interpersonal relationships (e.g., friendships) are constituted in part by the normative expectation that each party treats the other with respect.
9. If *A* has a normative expectation that *B* treat her with respect, then *A* will be disposed to resent *B* in the event that *B* fails to treat her with respect.
10. *A*'s disposition to resent *B* would be a fitting disposition to cultivate towards *B* only if *B* was, in principle, an apt target of resentment.
11. Thus, *A*'s normative expectation that *B* treat her with respect is a fitting expectation to make of *B* only if *B* is an apt target of resentment.
12. Thus, reciprocal interpersonal relationships are fitting forms of human engagement only if the parties to such relationships are apt targets of resentment.[9]

If this argument is sound, then it shows that being an apt target of resentment is a necessary condition on the legitimacy of ordinary reciprocal interpersonal relationships. Thus, the soundness of this argument secures an important point: viz., the agential capacities that would be sufficient for participation in ordinary interpersonal relationships must themselves be sufficient to render an agent an apt target of resentment. And because no one has ever argued that one must be the ultimate source of one's actions in order to legitimately stand in

[9] For more on this point, especially as it relates to friendships and other love relationships, see [author information].

relationships of this sort, there is no reason to think that one must be ultimate source of one's actions in order to be an apt target of resentment or other interpersonal attitudes.

But why should one accept this argument? Well, it seems to me that (8) and (10) are pretty difficult to deny. After all, (8) is simply a conceptual claim about reciprocal relationships—viz., if they are not governed by a standard of mutual respect, then they cannot be truly reciprocal. In the case of (10), its denial seems absurd. How could it be fitting to cultivate dispositions to engage with someone in a way that, as a matter of principle, cannot be deserved? There might be, of course, reasons to cultivate such dispositions—that doing so promotes good consequences, for example. But these are the *wrong kind of reasons* to cultivate dispositions to resent, feel indignant, etc. Only reasons grounded in the capacities that, at least in principle, render *B* a fitting or deserving target of resentment, et al. are the *right kind of reasons* for *A* to be disposed to resent *B*.[10]

Together with (9), these premises entail (11) and (12).[11] Since I take (8) and (10) to be relatively secure, this means that the real point of contention in the argument is (9). So what are the reasons to accept (9)? Well, following R. Jay Wallace, it might be that if one has a *normative* expectation about how another agent should act, then one is prepared to hold that agent responsible for failing to act in that way.[12] Accordingly, if you expect that I will show you respect then you are ready (i.e., disposed) to hold me responsible for failing to show you respect. But holding others responsible is not merely a matter of believing that they are responsible. One must also be emotionally exercised; one must resent or be indignant towards those who violate such expectations to really count as holding responsible.

---

[10] I'm grateful to an anonymous referee for helping me to clarify the details of this argument.
[11] To be clear: (12) is not meant to entail that friendships or love relationships couldn't be enjoyable, meaningful, or worth pursuing if it turns out that no one is an apt target of resentment. However, it does entail that such relationships would not be fitting forms of human engagement given the presuppositions of the attitudes entailed by the relationship. In other words, it entails that such relationships are defective by their own constitutive standards.
[12] R. Jay Wallace, *Responsibility and the Moral Sentiments*, Harvard University Press (1994).

Therefore if one genuinely expects respect of another, then one will be disposed to resent that person if they fail to show respect.

But here you might reasonably wonder why we should accept Wallace's claim that normative expectations for respect are constitutively tied to desert-entailing reactive attitudes like resentment or indignation.[13] No doubt, he's right to note that one has a normative expectation of another that they respect you only if you're disposed to be emotionally exercised by violations of this demand. But why must that entail that one is necessarily disposed to *resent* those who fail to show respect? Is it not possible to be disappointed, sad, and hurt by others' failure to respect you without being disposed to resent them? And if so, doesn't this mean that you can have the normative expectation of another that he or she treat you with respect without thereby being disposed to reactive emotions like resentment in the event that they fail to respect you?[14] If there is an affirmative answer to this question, then this calls into question the truth of (9).

In response to worries of this sort, several philosophers have recently attempted to show why emotions like sadness, disappointment, or hurt feelings cannot serve the same role as reactive emotions like resentment in the context of interpersonal relationships. Wallace himself claims that blaming emotions like resentment and indignation attest to the moral importance of reciprocal relations in a way that sadness and disappointment do not.[15] For unlike sadness or disappointment, resentment is available only to the person who has been shown disrespect. It's thus uniquely interpersonal as a form of response. And though

---

[13] Derk Pereboom (2001) and Tamler Sommers (2012) express worries about claims of this sort (though both Pereboom and Sommers are more focused on P.F. Strawson's statement of this general idea. Cf. Pereboom, *Living Without Free Will*, Cambridge: Cambridge University Press (2001); Sommers, *Relative Justice: Cultural Diversity, Free Will, and Moral Responsibility*, Princeton: Princeton University Press (2012).

[14] See Pereboom (2001) and especially, Derk Pereboom, *Free Will, Agency, and Meaning in Life*, New York: Oxford University Press (2014) for a defense of this claim.

[15] R. Jay Wallace, "Dispassionate Opprobrium: On Blame and the Reactive Sentiments," *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, eds. R. Jay Wallace, Rahul Kumar, and Samuel Freeman, Oxford: Oxford University Press (2011): 348-372.

indignation is similarly available to third parties it is not an *impersonal* response like sadness or disappointment, but one that imagines the wrongdoer as violating a norm of interpersonal respect that the *blamer* holds them to. In other words, reactive emotions like resentment and indignation are relational in a way that other emotions are not. It makes sense, then, that Wallace takes these emotions (and not sadness or disappointment) to be so tightly tied to the normative expectation of shared respect that is at the heart of reciprocal interpersonal relationships.[16]

One might try to further defend disappointment as a suitable analog to desert-entailing emotions like resentment and indignation by arguing more directly for the claim that disappointment is, like Wallace claims of resentment and indignation, constitutive of the stance of holding someone to a normative expectation. This is fairly plausible in a wide range of scenarios. Consider, for example, a teacher who expresses disappointment *at* her students when they fail to study for an exam. Even if the teacher is not prone to meet their laziness with resentment or indignation, it still seems that her proneness to disappointment is itself constitutive of her normative expectation. At the very least, it's certainly evidence that she takes herself to be holding them to that expectation. And if this is right, then perhaps it's possible to make sense of the sort of reciprocal interpersonal relationships that both Strawson and Wallace emphasize without appealing to the reactive emotions of resentment or indignation.

The problem with this, however, is that it simply assumes that disappointment is *not* a desert-entailing emotion. But if disappointment *were* a desert-entailing emotion, then this

[16] Seth Shabo makes a related claim about the connection between taking others' actions personally in a way that is required for love and taking their actions personally in a way that makes resentment a fitting response to disregard. For Shabo, however, the connection in question is apparently psychological. That is, one cannot, as a matter of human psychology, divorce one's regard for another as beloved from one's disposition to feel resentment towards that person in instances of disregard. For more, see Shabo, "Where Love and Resentment Meet: Strawson's Intrapersonal Defense of Compatibilism," *The Philosophical Review* 121.1 (2012): 95-124.

line of response would be of no aid to the moral responsibility skeptic, since it too would have to be excised from our lives. And although I'm not sure I have a decisive argument that disappointment is a desert-entailing emotion, I think there is some reason to think that it is.

To see this, let's note that just as one's resentment and indignation are sensitive to certain facts about the quality of another's will, so too, disappointment of the sort that we direct at others (I call this "disappointment *at*) is responsive to the reasons for which an agent acts. If the teacher who expresses disappointment *at* her students later discovers that they weren't lazy, but that they were all working jobs in the afternoons and evenings to help support their families, then it puts their failure to study in a new light. Their failure no longer displays an objectionable quality of will, but instead shows her students to be doing what is, in their situations, best. Naturally, the teacher might still feel some disappointment. But any disappointment she feels here will be disappointment *that* the world is unjust, disappointment *that* her students unfairly have to grow up too early, etc. However, this is not a form of disappointment that is directed *at* her students, and so it's not the sort of emotional response that could play the role of constituting her expectation. For once the teacher realizes that it's not fair to be disappointed at her students' poor performance—that her disappointment isn't *deserved*—she'll have to give up on really holding them to the expectation in question.

This suggests a more general point about disappointment of the sort that can be directed *at* others. When others express disappointment *at* our actions, it seems that we can complain about their response is precisely the sorts of situations in which resentment's excusing and exempting conditions arise. And the best explanation for this is simply that like resentment and indignation, disappointment *at* is a desert-entailing emotion—i.e., it is a fitting response to another's action or character only if that person is morally responsible for

their action or character. In other words, the emotional responses that can plausibly serve as functional analogs of resentment or indignation—i.e., the emotional responses that can plausibly serve to ground normative expectations of the sort that are internal to reciprocal interpersonal relationships—will also be desert-entailing emotions. This means that *even if* it's true that disappointment *at* can serve to ground normative expectations of the sort that are required for interpersonal relationships, it only can do so because it, like resentment and indignation, presupposes that those who violate such expectations *deserve* to be held responsible.[17]

As I said above, however, this argument is not decisive, but it does apparently shift the explanatory burden back on the skeptic. In any case, one can supplement the support of (9) that I've offered here by further articulating the axiological basis for the claim that desert-entailing emotions like resentment and indignation (and maybe even disappointment *at*) cannot be excised from our relationships with others.

There is a notable connection here between what I say here about normative expectations of respect and what many have said about *valuing* more generally. To *value* something is, of course, more than simply *judging* it to be *valuable*. For example, I might think that it is good that my city has a vibrant arts scene, and yet not actually value that myself. Valuing, it seems, requires more than mere belief; it requires *inter alia* that I be emotionally engaged with the object of value in a way that manifests a concern to protect and defend it. Attitudes like sadness and disappointment *that* do not, and indeed *cannot*, manifest this

[17] I actually think this sort of argument will generalize to any form of blame or holding responsible that might serve as a basis of normative expectations. That is, any putative analog to the reactive emotions, whether it's disappointment *at*, or the kinds of blame that George Sher (2006) and T.M. Scanlon (2008) have offered, will be subject to the same set of excusing and exempting conditions as reactive emotions. And the best explanation for this, I claim, is that these other forms of blame are also desert-entailing. Neither Sher nor Scanlon make this point (and Scanlon himself rejects it), but how else to explain the fact that reactive emotions are governed by norms that are isomorphic to the norms that govern these forms of blame? For more, see George Sher, *In Praise of Blame*, Oxford: Oxford University Press (2006) and T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame*, Cambridge, MA: Harvard University Press (2008).

concern, since these attitudes and the behaviors they give do not serve to *protest* attacks on the object of value so much as to *lament* it. By contrast, the emotions of resentment and indignation (and possibly, disappointment *at*) do serve to protest harms knowingly done to the object of value.[18] Of course, one won't fail to value something simply because one is not disposed to respond to attacks on that thing's value with resentment or indignation, but there does seem to be something deficient about one's attitude towards the thing *qua* act of valuing in such cases.

Since we value (or should value) our fellow persons, we are concerned (or should be concerned) that they are treated with respect. In fact, it's for this very reason that we have normative expectations that they be treated with respect. So too, we value ourselves, and so we have the normative expectation of others that they will regard us with respect. It's for this reason that violations of the expectation of respect cannot exclusively be met with sadness or disappointment. True, it is lamentable, and in some cases even tragic, that we act in ways that harm and disrespect others. It's for this reason sadness and disappointment *that* aren't wholly out of place in the event of wrongdoing. But to value others fully, wrongdoing must also to be condemned and protested, and it is through reactive emotions that this is done. It's for this reason that I'm inclined to think that (9) is true, such that if one is not disposed (even in the slightest) to attitudes like resentment or indignation in the event that another acts wrongly, then one does not really hold that person to the normative expectation that he or she treats others with respect.[19]

---

[18] See Pamela Hieronymi, "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62.3 (2001): 529-555; Christopher Evan Franklin, "Valuing Blame," *Blame: Its Nature and Norms*, eds. D. Justin Coates and Neal A. Tognazzini, Oxford: Oxford University Press (2013): 207-223.

[19] It's worth noting that this is consistent with thinking that one has all things considered reason to refrain from ever expressing one's resentment of others. In fact, if generosity, meekness, and forgivingness really are virtues (as many have previously held), then we might want to regard others with more equanimity even though we are disposed to resent them for their transgressions.

Admittedly, this defense of (9) is quick, but I think it's at least reasonable to accept the premise on this basis. As a result, I think it's reasonable to think that the above argument is sound. However, even if one accepts the soundness of this argument, it's nevertheless reasonable to doubt that the agential capacities required for the legitimacy of reciprocal interpersonal relationships would suffice to render an agent an apt target of eternal torment, since damnation is a form of punishment that is not internal to ordinary interpersonal relationships. In other words, because interpersonal relationships do not themselves presuppose the existence of God (or that he would ever damn wrongdoers), the capacities required to participate in such relationships would be seemingly independent of the capacities that might underwrite just damnation.

This point can be clarified by considering sourcehood. As I argued earlier, it's plausible that sourcehood is a requirement on being an apt target of eternal damnation. To be ultimately morally responsible, then, I must be the ultimate causal source of my actions. But must I be the ultimate causal source of my actions to be friends with someone? Or do I need to be the ultimate causal source of my actions to stand in any of the other ordinary interpersonal relationships that we hold so dear? I can't see why you'd think that, since the agential capacities relevant to these sorts of relationships seem to be those that underwrite our ability to understand and respond to the norms of these relationships and the ability to engage in joint activity. Thus, if it were to turn out that we were not apt targets of eternal torment, it's not clear why we should thereby doubt the legitimacy of our interpersonal relationships. For even if we are not ultimately morally responsible—and following Galen Strawson, I've already argued that we have good reason to doubt that we are—we might nevertheless be apt targets of resentment, since we might nevertheless be capable of

participating in interpersonal relationships and attitudes like resentment are internal to such relationships.

No doubt, if we are not ultimately morally responsible then we lack an agential capacity that is of some value—*who wouldn't want to be a prime mover?* But the fact that insofar as we are capable of participating in legitimate interpersonal relationships we are thereby apt targets of attitudes like resentment is itself nothing to scoff at. And sensibly, following P. F. Strawson, a number of theorists have claimed that this is sufficient for morally responsible agency. For example, R. Jay Wallace has claimed that an agent instantiates the property of being morally responsible for their actions if and only if it is appropriate for others to hold her responsible (where this invariably invokes the attitudes of resentment or indignation).[20] Even more plausibly, John Martin Fischer and Mark Ravizza have suggested that an agent is morally responsible insofar as she is an apt target of reactive attitudes like resentment.[21] In other words, the property of being an apt target of interpersonal attitudes like resentment just is the property of being morally responsible (in some sense) for one's actions.

Of course it is very doubtful that instantiating the property of being morally responsible in *this* sense renders an agent deserving of eternal torment. This is because the property in question is connected with ordinary interpersonal relationships, and within our interpersonal relationships, it's not particularly important that we're the ultimate causal sources of our actions, though it's very important that we're capable of acting in ways that adequately respect others' status as equals.[22] As a result, of course, the notion of moral

---

[20] See Wallace, (1994).

[21] I say "more plausibly" not because I think Wallace's view is implausible but because Fischer and Ravizza's view is not subject to an objection that Angela Smith has recently leveled against Wallace's view. For more, see John Martin Fischer and Mark Ravizza, *Responsibility and Control*, Cambridge University Press (1998); see also, Angela Smith, "On Being Responsible and Holding Responsible," *Journal of Ethics*, 11.4 (2007): 465-484.

[22] More needs to be said to provide a full account of this capacity. But plausibly, to be capable of this form of respect, an agent would need to be appropriately sensitive to reasons for respecting others. However we analyze this capacity thought, it is what serves as the basis for our participation in ordinary reciprocal

responsibility tied to being an apt target of reactive attitudes like resentment is more metaphysically modest than ultimate moral responsibility. Consequently, although being *modestly* morally responsible will open an agent up to the kinds of blame that are central to interpersonal relationships like the reactive attitudes (and paradigmatic expressions of these attitudes), instantiating this property does not appear to, by itself, suffice for more severe sanctions (including eternal torment).

Nevertheless, this doesn't mean we should reject this notion of modest moral responsibility out of hand as insignificant or as not the "real McCoy" or "genuine article" with respect to morally responsible agency. No doubt, much of the historical motivation for theorizing about moral responsibility did issue from concerns about the fairness of divine punishment. But this isn't the only reason we should care about being morally responsible. We might think that issues of blame, which arise primarily in the context of interpersonal relationships, are also important ones to theorize about. And given the difference between the relative severity of enduring hell and being blamed, it shouldn't be too surprising that the capacities that underwrite the latter are more modest than the capacities that underwrite the latter. Thus, though the resultant theory of responsibility doesn't require such metaphysically grandiose capacities as the ability to be a prime mover, it's still normatively important.

4. The Basic Argument Revisited

Therefore, with this brief statement of modest moral responsibility in mind, I now want to return to Galen Strawson's Basic Argument. Earlier, I suggested that it might be sound—as long as we take seriously the idea that the notion of "moral responsibility" that is employed in premise (1) is the notion of ultimate moral responsibility. But I hope that we are now in a

interpersonal relationships. For richer accounts of what such a capacity would look like, turn to the usual suspects: Wallace (1994), Fischer and Ravizza (1998), and Darwall (2006).

position to see that premise (1), when disambiguated as a principle putatively governing modest moral responsibility, is false.

If being modestly morally responsible is a property we instantiate in virtue of those agential capacities that allow us to participate in reciprocal interpersonal relationships, then there is no reason to think that an agent would be modestly morally responsible for her decision to *x* only if she is morally responsible (in any sense) for the state of mind that issues in her *x*-ing. To see this, consider that you should still feel guilty for letting a friend down, even if it turns out that at the some point, the etiology of your action was due to factors for which you are not the source. This is because guilt is made fitting by wrongful disregard, and the mere fact that the causal history of your action traces through factors outside of your control does not obviate the fact that you were capable of appreciating the reasons you had for upholding the norms of friendship or the fact that at the time, you had no interest in doing so. Thus, guilt seems fitting on your part. And consequently, it seems that you could be modestly morally responsible for your action even if you were not modestly morally responsible for the state(s) of mind that issued in that action.

A version of the Basic Argument that invokes modest moral responsibility is therefore unsound. And because we should care about modest moral responsibility at least as much as we care about ultimate moral responsibility (I have to admit, I am much more concerned that I instantiate the former property than I am the latter, but on this point I think reasonable people can disagree), the Basic Argument gives us no reason to doubt that we are morally responsible.

5. Conclusion

Galen Strawson's skeptical argument might be sound on some readings, but it gives us no reason to doubt our status as modestly morally responsible agents. But so what? How should this affect our theorizing about freedom and responsibility? As a first pass at answering these questions, I want to conclude by taking a step back and situating modest moral responsibility in its wider context.

First, it's important to note that nothing I've said here prejudges the debate between compatibilism and incompatibilism. Although I am a compatibilist about such issues, one could think that the ability to do otherwise (holding fixed the past and the laws) is necessary for reciprocal interpersonal relationships, since without it, we would lack the relevant ability to recognize and *respond* to reasons. This would result in an incompatibilist theory of modest moral responsibility without leading one to equivocate between modest and ultimate varieties of responsibility. Of course, whether interpersonal relationships require the ability to do otherwise (holding fixed the past and the laws) is another matter. So whether the property of being modestly morally responsible can be instantiated in a causally determined world is an issue that must be resolved. But we cannot resolve *this* issue unless we're careful not to run together modest and ultimate varieties of responsibility, as is too often the case.

Second, it's important to see that I haven't changed the subject in any objectionable way. This might not be readily apparent from the initial argument. Indeed, one might worry that part of what motivates Strawson to frame ultimate moral responsibility in terms of God's justice is that fixes the issue in a way that excludes consequentialist or contractualist considerations. However, the shift in emphasis to interpersonal relationships that motivates modest moral responsibility doesn't clearly exclude these considerations. But such considerations are irrelevant to questions of desert. So, what I'm really talking about when

I'm talking about modest moral responsibility is a non-desert entailing conception of responsibility, and skeptics like Strawson have no reason to deny that we might be morally responsible in non-desert entailing ways.[23]

In response, I want to first note that I have changed the subject, though not, I think, in any *objectionable* way, since our lives together give us ample reason to care about modest moral responsibility. Moreover, it's important to remember that the property of being an apt target of reactive attitudes like resentment, which figures prominently in (10) – (12) of the above argument, *is* meant to exclude consequentialist or contractualist considerations. In fact, the reason I prefer talk of 'aptness' to 'appropriateness' is that the latter suggests that consequentialist or contractualist considerations *do* matter, since part of what makes it appropriate for me to blame you is that it will do some good. But whatever good my blame might do, this isn't relevant to why you are an *apt target* of my blame. Talk of being an apt target, then, is meant to track our pre-theoretical intuitions about desert, and I don't think anything about the way I explicate modest moral responsibility requires that it is not related to desert in precisely the same way that many have taken ultimate moral responsibility to be related to desert. This means that the skeptic cannot accept that agents are modestly morally responsible and continue to maintain that no one is responsible in the desert entailing sense. In other words, in introducing modest moral responsibility, I have changed the subject from divine retribution to human emotions, but I have not changed the subject from desert entailing forms of responsibility to non-desert entailing forms of responsibility.[24]

---

[23] I'm grateful to an anonymous referee for pressing me on this point.

[24] A related worry (which was also helpfully pressed by an anonymous referee) concerns whether our negative reactive emotions are retributive in the same way that divine punishment is thought to be retributive. If not, then again, one might worry that I have changed the subject in an objectionable way. I think that what I say in response to the worry about desert will also apply here. Additionally, I want to also point out that following P. F. Strawson, many have taken the negative reactive emotions to have a retributive element. Strawson himself seems to think that part of what it is to resent is to be willing to acquiesce in the suffering of the wrongdoer. But others, including those who argue that resentment, et al. are fundamentally misguided also take them to

Finally, I want to discuss the connection between modest moral responsibility and criminal punishment. I've argued that being modestly morally responsible for *x*-ing (where *x*-ing is wrong) can make negative reactive attitudes and perhaps certain forms of institutional discipline (being written up by a supervisor for insubordination) deserved. But, whether one is deserving of these forms of blame is a relatively low stakes matter, especially when we remind ourselves of Galen Strawson's invocation of divine punishment. What about cases in which the stakes are higher than interpersonal blame but lower than divine punishment—e.g., cases in which an agent is liable to be held *criminally* responsible for his or her action?

I think it is an open question whether the kind of control required for modest moral responsibility will underwrite the possibility that an agent might deserve long-term imprisonment or execution. (This holds even if we are assuming that the criminal justice system that is enforcing such punishments is, unlike our own, a just one.) I tend to think not, and so the skeptical force of Galen Strawson's argument is perhaps stronger than I have given it credit: maybe it's not just divine punishment that no one deserves but also (sufficiently severe) criminal punishment. And if so, recognizing that it's modest and not ultimate moral responsibility that underwrites many of our informal blaming practices will not save us from needing to make significant revisions to (some of) our responsibility practices. Thus, as an anti-skeptical response, my appeal to modest moral responsibility might be significantly weakened.

As a final reply, I will happily concede the point: being modestly morally responsible will probably not suffice to render us deserving of just any response to our wrongdoing that

---

have this retributive element. See, e.g., Martha Nussbaum, *Anger and Forgiveness: Resentment, Generosity, and Justice*, Oxford University Press (2016).

falls short of damnation. But it does render us apt target of forms of blame that are internal to the human relationships that give our lives meaning. And this is no modest result.