

The Physiognomy of Responsibility

JOHN MARTIN FISCHER

University of California, Riverside

NEAL A. TOGNAZZINI

The College of William & Mary

Our aim in this paper is to put the concept of moral responsibility under a microscope. At the lowest level of magnification, it appears unified. But Gary Watson has taught us that if we zoom in, we will find that moral responsibility has two faces: attributability and accountability.¹ It is one thing to talk about the connection the agent has with her action; it is quite another to talk about the potential interaction the agent might have with her moral community.²

It turns out, though, that the faces of moral responsibility can themselves be viewed under an even higher level of magnification. If moral responsibility has two faces, then our aim in this paper is to examine their features. To do so reveals subtle distinctions in our concept of moral responsibility and its interaction with surrounding issues that, we argue, can help illuminate various debates in the literature.³

¹ See Watson 1996. It's unclear to us, however, whether our use of these terms throughout this paper matches up precisely with Watson's usage. In any case, although Watson's distinction is a source of inspiration for this paper, we are not attempting exegesis.

² For an excellent discussion of the importance of keeping these two faces apart, see Smith 2007. R. Jay Wallace argues, however, that these two faces are importantly connected in that someone *is* morally responsible just in case it is appropriate to *hold* that person morally responsible. See Wallace 1994, especially chapter 4.

³ Manuel Vargas has pointed out to us (in personal correspondence) that what we go on to investigate may not all belong to the concept of moral responsibility proper, but instead to other related concepts. That may well be right. We merely use the word 'concept' here to point out that we are not investigating any particular *conception*, or *conditions for the application*, of the concept of moral responsibility. Rather, we are investigating the concept itself (and surrounding issues). For more on the distinction between concept and conception as it applies to moral responsibility, see chapter 1 of Fischer and Ravizza 1998.

1. **Attributability: Aretaic and Reactive**

We begin our examination with attributability because there is a sense in which questions about attributability are analytically prior to questions about accountability. We shall then proceed stepwise through each analytical “stage” or “step” on our way to accountability of the strongest sort. It is important to keep in mind, however, that our use of the terms ‘step’ and ‘stage’ is not meant to imply anything about *degrees* of moral responsibility. These are *analytical* or *conceptual* stages, which may or may not have any correlation with *how much* moral responsibility an agent possesses, if such an idea even makes sense. An agent who is at, say, the 15th step is not necessarily *more* morally responsible than an agent who is at, say, the third step. Additionally, it is important to note that these are analytical moments, not necessarily developmental or chronological moments or stages.

Since our investigation involves working through the various analytical stages of moral responsibility, perhaps the most straightforward way to proceed is to imagine that we have been given the task of determining whether someone who is alleged to have done something morally wrong is indeed morally responsible for the action in question. How shall we figure it out? Let the agent be ‘Sam’, and the alleged moral wrong in question be ‘A’. What would we need to know about Sam, A, or the relationship between the two that would help us figure out whether Sam is morally responsible for doing A? A natural place to begin our investigation is with what might be considered a “precondition” for moral responsibility:

Question 1: Did Sam do A?

Clearly, the first thing we want to know is whether Sam even performed the action in question. There are two situations in which question 1 might receive a negative answer. First, if Sam has an airtight alibi according to which she had absolutely nothing to do with the morally wrong deed, then Sam clearly isn’t morally responsible for the deed in question. We should let Sam go and look for other suspects. Second, we would answer question 1 negatively were we to find out that although Sam was involved in the alleged morally wrong behavior, the behavior wasn’t an action of hers. That is, perhaps the behavior was the result of an uncontrollable muscle spasm, or maybe someone else was controlling Sam’s limbs in such a way that the event we have identified as A resulted. If Sam has an alibi, then she gets off the hook because *she* didn’t do anything wrong; if A wasn’t actually an action of

Sam's, then Sam gets off the hook because she didn't *do* anything wrong.

Suppose, however, that we discover that Sam did indeed perform A, and thus the precondition to moral responsibility is satisfied. Now we can start to ask about the senses in which the action is attributable to Sam. In particular, we want to know about the capacities that Sam exercised in A-ing:

Question 2: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of *aretaic* appraisal?

The two emphasized terms need to be explained. Let's start with 'aretaic'. Watson uses this term to describe the face of moral responsibility that is concerned with attributability, and he picked it because aretaic appraisals "concern the agent's excellences and faults – or virtues and vices – as manifested in thought and action."⁴ Importantly, one need not think that conduct is *morally wrong* in order to make a negative aretaic appraisal of the conduct. One need only judge that the conduct is faulty in some way that reflects poorly on the agent. Watson gives the following example of faulty conduct that may not be morally wrong:

If someone betrays her ideals by choosing a dull but secure occupation in favor of a riskier but potentially more enriching one, or endangers something of deep importance to her life for trivial ends (by sleeping too little and drinking too much before important performances, for example), then she has acted badly – cowardly, self-indulgently, at least unwisely.⁵

In this case, we can conclude that the agent is open to aretaic appraisal: the conduct is faulty in a way that reflects poorly on the agent. What we *can't* conclude, however, is that this cowardly person is open to blame from the members of her moral community.⁶ It may be that openness to blame requires something more. When we ask, then, whether Sam is open to aretaic appraisal on the basis of her A-ing, we

⁴ Watson 1996, as reprinted in Watson 2004, p. 266.

⁵ Watson 1996, as reprinted in Watson 2004, pp. 266-267.

⁶ We use the phrase 'open to' in a way that makes it roughly synonymous with 'is a sensible target of', rather than in the stronger sense of 'is justifiably subject to'. Thus when we say that an agent is open to aretaic appraisal, we simply mean that such appraisal would not be a category mistake. Whether such appraisal is justifiably expressed to the agent, for instance, is not a matter on which we here take a stand.

are not (yet) asking about whether she is open to blame from her moral community.⁷

Nomy Arpaly is another theorist who appears to take seriously the concept of aretaic appraisal. On her view, “blame is analogous to holding someone to be a bad businessman or a lousy artist.”⁸ Although we prefer to reserve the term ‘blame’ for talking about accountability (see note 7), it is clear that Arpaly is here talking about some sort of aretaic appraisal. To say that someone is a bad businessman is merely to say that his conduct as a businessman is faulty in some way, that it doesn’t achieve excellence. Similarly for the lousy artist. In making these judgments, we are *attributing* the faulty business dealings and the faulty art activities to the agent in a way that reflects poorly on the agent. We are thus engaging in aretaic appraisal of the agent.

By asking whether the agent is open to, or is a ‘sensible target’ of, aretaic appraisal, we are asking whether the agent exercised the capacities required to make the agent *the sort of creature whom it might make sense* to appraise aretaically. That is, we need to know whether making an aretaic appraisal would be a category mistake, as it surely would if the agent in question were, say, a dog.⁹ Once we have an affirmative

⁷ One might think that even to make an aretaic assessment is to blame, in a sense. While we are inclined to disagree, we need not take any substantive position on the concept of blame at this point. Perhaps there is a sort of blame associated with attributability. However, we would urge the reader to keep in mind the distinction between responding to wrongdoing with a judgment, on the one hand, and responding with some sort of overt interaction, on the other. Whether one or both of these responses count as *blame* is less important to us at the moment than the fact that aretaic appraisal seems most naturally to fit into the *judgment* category. Our use of the word ‘blame’ in the text is meant to indicate something more robust than mere judgment, but we certainly recognize that some will object to our view that blame involves something more than a type of judgment. See, for instance, Hieronymi 2004 and Sher 2006.

⁸ Arpaly 2003, p. 173.

⁹ We have not said much about the nature of aretaic appraisal at this point, so one might legitimately wonder why we say that dogs are not sensible targets for it, especially if it is merely a matter of excellences and faults, as the etymology suggests. Here we would appeal to Watson’s own characterization of aretaic appraisal, which he says is concerned not only with skill but also with “one’s purposes, ends, choices, concerns, cares, attachments, and commitments”. For Watson, aretaic appraisal “presupposes moral capacity, the capacity for adopting and pursuing ends” (Watson 2004, pp. 287-288). Although dogs can have excellences and faults in some sense, the sense required for aretaic appraisal is more robust. So, for example, a dog can be “vicious”, but its viciousness is not specifically moral viciousness; a vicious dog is (considerably) more aggressive and more inclined to bite than the average dog, but it surely does not intend to cause pain or suffering. This may distinguish a dog from certain psychopaths, who can indeed have specifically moral intentions (although they may not be able to grasp that there are moral reasons to refrain from the behavior in question). We thank an anonymous referee for pushing us to be clear about this issue.

answer to question 2, we know *something* about attributability. We know that Sam's A-ing is attributable to her in the aretaic sense: the capacities she exercised in A-ing make her a sensible target for aretaic judgments about her conduct. It wouldn't be a category mistake to judge, for instance, that her conduct is faulty in a way that reflects poorly on her as an agent.

But we maintain that openness to aretaic appraisal is not the only sort of attributability. Although Watson certainly identified an important sense of attributability by pointing to aretaic appraisal, there is at least one more sort of attributability, and it is this sort that is importantly connected with the second (accountability) face of moral responsibility.

It is natural to suppose that moral responsibility is intimately connected with what P.F. Strawson called 'reactive attitudes', of which guilt, resentment, and indignation are uncontested examples.¹⁰ As we will see below, these reactive attitudes form the basis from which to build up the *accountability* face of moral responsibility. But before we get to the accountability face, we need to know whether the agent is even in the accountability game. That is, just as question 2 told us that the agent exercised the capacities needed to be open to aretaic appraisal, our next question will tell us whether the agent exercised the capacities needed to be open to the reactive attitudes:

Question 3: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of the reactive attitudes?¹¹

¹⁰ See Strawson 1962. There are other plausible ways of understanding moral responsibility that do not make essential reference to the reactive attitudes. For example, Michael Zimmerman is attracted to a "ledger-view" of moral responsibility according to which ascriptions of moral responsibility amount to claims about what sorts of "debits" and "credits" one has accrued in one's moral ledger. This is admittedly metaphorical, and Zimmerman attempts to spell out the metaphors a bit more in his 1988, especially pp. 38-39. For the purposes of this paper, we merely take as a working hypothesis an account of moral responsibility that involves the reactive attitudes.

¹¹ John Martin Fischer and Mark Ravizza have said that a positive answer to question 3 tells us that Sam is morally responsible for A-ing. That is, they use the term 'moral responsibility' in this way: "Someone is morally responsible *for a particular bit of behavior* (or perhaps a trait of character) to the extent that he is an appropriate candidate for at least some of the reactive attitudes on the basis of that behavior (or trait of character)" (Fischer and Ravizza 1998, pp. 6-7). When they say 'appropriate candidate', they mean what we mean by 'sensible target': it would not be a category mistake to apply the reactive attitudes to the agent in question on the basis of her behavior. But we bring up Fischer and Ravizza here only to orient ourselves with respect to the literature; we do not want to prejudge any question about how the term 'moral responsibility' ought to be used. Indeed, part of our project here is to argue that there are many equally appropriate ways to use the term that need to be distinguished.

To fix terms, let us say that an affirmative answer to question 3 tells us that Sam is the sort of being who is open to *reactive* appraisal on the basis of A-ing.¹² Now, you might think that a positive answer to question 2 automatically generates a positive answer to question 3. That is, you might think that if Sam is open to aretaic appraisal on the basis of A-ing, then she is also a sensible target of the reactive attitudes on the basis of A-ing. After all, those creatures that are not sensible targets of the reactive attitudes – like dogs and cats, for example – seem also not to be open to aretaic appraisal. And it may well be that when they act most normal human adults typically exercise the capacities that make them sensible targets for both aretaic and reactive appraisal. Nevertheless, the two stages may come apart. To see how, consider the vexing case of the psychopath.

There is much dispute about whether psychopaths possess the capacities needed to be morally responsible agents. We will not take a stand on that issue here, but we do note that any answer will have to specify just what *sort* of moral responsibility is at issue. When we wonder about a psychopath's moral responsibility, what exactly is the issue? If it is whether we can legitimately praise and blame him, then this is a question about accountability (as we will see in more detail below). If the issue is whether he is the sort of creature whom it makes sense to target with the reactive attitudes, then this is a question about reactive attributability. And finally, if the issue is whether he is the sort of creature who might be appraised aretaically, then this is a question about aretaic attributability. Importantly, these last two issues may come apart if, for instance, the ability to respond to specifically *moral* reasons is necessary for anyone to count as a sensible target for the reactive attitudes. Psychopaths will then plausibly fail to satisfy the conditions of reactive attributability (since they are commonly thought to lack the capacity to respond to moral reasons), though not those for aretaic attributability. One can still be cowardly and self-indulgent even if one cannot respond to moral reasons.¹³ Thus, it is coherent to suppose that the behavior of psychopaths is attributable to them in an

¹² We use the term 'reactive' here in a stipulative sense merely to point to the connection that question 3 has with the reactive attitudes. Aretaic appraisal comprises a judgment about faulty conduct, but does not necessarily involve any judgment about what sorts of responses might be appropriate from members of the agent's moral community. Reactive appraisal, in our sense, takes this further step.

¹³ This is not to say, of course, that it would be useful to point these faults out to the psychopath, who may, after all, not even understand what it is for behavior to be self-indulgent. It would remain self-indulgent behavior, however, despite the futility of trying to get the psychopath to understand it as such. Thanks here to an anonymous referee.

aretaic sense, even if it is not attributable to them in a *reactive* sense. For this reason, we have distinguished between questions 2 and 3.¹⁴

We have seen that there are at least two kinds of attributability: aretaic and reactive. While it is plausible to suppose that the actions of most agents are attributable to them in either both senses or neither, this supposition may fail in cases like that of the psychopath. Nevertheless, once we have positive answers to questions 2 and 3, we know everything there is to know about the *attributability* face of moral responsibility.¹⁵ You might think that we can now move immediately to the *accountability* face. However, it turns out that there are at least two more questions to be answered before we get to the issue of accountability. There is an analytical gap, so to speak, between the two faces of moral responsibility. The next two questions fill that gap.

2. The Space Between

Another way to talk about reactive attributability is to talk in terms of *exemption*. A negative answer to question 3 – that is, a judgment that an agent is not a sensible target of the reactive attitudes and thus is not open to reactive appraisal – amounts to the claim that the agent ought to be *exempted* from the practices that we engage in when we hold each other accountable. As we have seen, however, it may be that not all exempted agents are relevantly similar. Psychopaths may be exempted because they cannot respond to moral reasons, but this does not mean that their actions are not attributable to them in an aretaic sense. Dogs are presumably open to neither aretaic nor reactive appraisal. Once we have an affirmative answer to question 3, then, we know that Sam is not to be exempted from our practices of holding each other accountable. Thus, judgments of reactive attributability are judgments of non-exemption.

Mere non-exemption, however, is not sufficient to take us to the accountability face of moral responsibility. This is because there is more than one way to get off the proverbial hook associated with accountability. Exemption is one way; justification is another. The next question we need to ask, then, is whether Sam had any *justification* for A-ing. We take it that this amounts to the following question:

Question 4: Was Sam's A-ing morally wrong?

To judge that someone's behavior was justified is, in our sense of the term, to judge that the behavior did not violate any moral norms after

¹⁴ For an interesting discussion of the significance of asking whether psychopaths are open to aretaic appraisal, see Levy 2007.

¹⁵ This claim is qualified in section 8 below.

all. Suppose you are attempting to kill a very large and dangerous spider, which happens to be crawling on Bill's foot. In order to step on the spider, you must step on Bill's foot as well. It may be that you exercised all the capacities required to make you a sensible target of the reactive attitudes when you stepped on Bill's foot, but it is plausible to suppose that your action was justified and hence that you haven't done anything wrong.¹⁶

Were we to find out that although Sam's A-ing was attributable to her in a reactive sense, she had a justification for A-ing, it would not be right for us to hold Sam accountable for A-ing.¹⁷ This is one way in which there is a space between reactive attributability and accountability. But it turns out that there is another gap, as well. Let's suppose now that question 4 has also been answered in the affirmative: Sam's A-ing is attributable to her in a reactive sense, and she has no justification. It remains to ask whether Sam should be *excused* from accountability for A-ing. So far, we have seen two ways to be *off* the hook: exemption and justification. But it may still be that someone should be excused even when a morally wrong action is attributable to that person in a reactive sense.¹⁸ This will constitute our fifth question:

Question 5: Is there no other factor that excuses Sam for A-ing?

Some may be skeptical about our distinction between questions 4 and 5, so we should say something about why we think they should be kept apart.

Suppose that a mother is faced with an awful choice: either she can save her own child from drowning, or she can save five other children

¹⁶ This example is based on one found in Scanlon 1986, p. 162. Thanks to Angela Smith for pointing us to this example.

¹⁷ An anonymous referee points us to another possible sense of the term 'accountability' according to which Sam would still be accountable for his behavior even if he had a justification. In this sense, to say that Sam is accountable is merely to say that it is legitimate to demand that Sam either tell us why his behavior was justified or else apologize. To be clear, when we say that it is inappropriate to hold Sam accountable for A-ing if he has a justification for doing so, we do not mean to rule out the legitimacy of this sort of demand. Perhaps we could mark this by saying that although Sam is still *answerable* for his behavior, he is not *accountable* for it, but in any case when we talk about holding someone accountable, we mean more than merely demanding an explanation. There are sticky issues here about the relationship between holding someone accountable, holding someone responsible, and blaming someone (among other concepts), but we merely flag the stickiness without resolving it.

¹⁸ Talk of excuse can sometimes lead to confusion because of the difference between "excused from responsibility" on the one hand, and "excused from accountability", on the other. Here we are using 'excuse' in the latter way.

from drowning. Perhaps the children are swimming in a lake and the mother is driving a boat but cannot get to all six of the children in time. She must choose. Suppose, further, that she chooses to save her own child instead of the five other children. Such a decision is surely understandable, but even more than that, it is at least arguable that it would be unjustified, even in principle, for anyone to target this mother with any negative reactive attitudes. This will presumably be because of the enormous difficulty involved in doing the right thing. And yet – here’s the important point – it’s plausible to suppose saving her own child is a morally wrong action that is attributable to her in a reactive sense. The story could no doubt be filled out in a plausible way to make the mother’s action morally wrong on a consequentialist theory, and it seems plausible to suppose that her action would be considered morally wrong on most deontological theories, as well. (Kantian theories would presumably deplore the mother’s partiality to her own child.) If this is right, it shows that positive answers to questions 3 and 4 do not automatically generate a positive answer to question 5. This mother is not exempted from our practices (question 3), and she has done wrong (question 4), but it seems that she should be excused (question 5). (Note that, for our purposes here, it is not essential that it be uncontroversial that the mother has both done the wrong thing and is not held accountable; all that is required is that it be a plausible view that can’t clearly be ruled out.)

Perhaps another example will prove useful here. Michael Zimmerman also argues for the distinction between questions 4 and 5 by presenting the following case:

Suppose Jones comes upon a car accident; the driver is unconscious. Being a member of the “TV generation”, Jones expects the car to explode at any moment, and so he rushes to the driver and drags him clear of the wreck. The result: the driver is paralyzed for life (whereas he would not have been if Jones had left him where he was), and the car does not explode. Now, did Jones do wrong?¹⁹

Zimmerman goes on to analyze this as a case where Jones has indeed done something wrong, but he nevertheless ought to be excused from our practices of accountability.²⁰ Why? In this case, the reason

¹⁹ Zimmerman 1988, p. 41.

²⁰ In Zimmerman’s terminology, although Jones has done something wrong, he is not *culpable*. That is, a judgment that his moral ledger has been tarnished by this action would be an incorrect judgment. Or, what seems to amount to the same thing, a judgment that he is blameworthy would be an incorrect judgment.

presumably has something to do with what Jones believed when he was pulling the driver from the car. Jones of course was not aware (and, we can suppose, was not culpably ignorant) of the fact that moving the driver would cause permanent paralysis, and hence Jones did not believe that he was doing something wrong. But of course believing that what you are doing is the right thing to do does not make it the right thing to do. In this case, Zimmerman thinks, and we agree, that it's plausible to suppose that Jones has done something morally wrong, even though his beliefs at the time help to excuse him from accountability.²¹ Of course, it is often difficult in practice to determine whether a particular action should count as an excused wrong or as not wrong at all, but we maintain that the analytical distinction is nevertheless real and significant.²²

Questions 4 and 5, we submit, fill the gap between reactive attributability and accountability. Once we have answered the first five questions in the affirmative, we can turn to asking about whether holding Sam accountable for A-ing is justified.

3. Accountability: Reactive Attitudes and Beyond

Once we know that Sam is not to be exempted or excused for a morally wrong action that she performed, we automatically know that we should answer our first question about accountability in the affirmative:

Question 6: Is it justified, in the circumstances, to target Sam with any of the reactive attitudes?

It's one thing to ask whether *it makes sense* to target Sam with the reactive attitudes on the basis of her A-ing (question 3), and quite another thing to ask whether, in the particular circumstances, anyone is in principle *justified* in targeting Sam with any of the reactive

²¹ One way to account for this example from Zimmerman is to revise question 4 along the following lines: "Was Sam's A-ing morally wrong, and did she know it was?" As we will see in more detail below, this is what Derk Pereboom does in an attempt to eliminate the need for question 5. But the case of the mother's decision to save her own child cannot be so easily accommodated, and in any case the addition of the epistemic component still shows that questions 4 and 5, as we have formulated them, are analytically distinct.

²² Aristotle appears to agree that sometimes we ought to excuse wrongdoers from accountability: "In some cases there is no praise, but there is pardon, whenever someone does a wrong action because of conditions of a sort that overstrain human nature, and that no one would endure." (*Nicomachean Ethics* 1110a24-1110a26, trans. Irwin and Fine 1995). Thanks to Angela Smith for this reference.

attitudes (question 6).²³ As we have seen, to give an affirmative answer to the latter question requires giving affirmative answers to the questions about justification and excuse as well.²⁴

With question 6, we have moved to the accountability face of moral responsibility. Inasmuch as the reactive attitudes are at least incipiently communicative, it seems plausible to suppose that to target an agent with the reactive attitudes on the basis of an action is to *hold* the agent accountable for that action. But of course there are various ways of holding someone accountable, from the mild to the extreme. Question 6 asks whether it is in principle justified to hold Sam accountable in the mildest sense: by targeting her with the reactive attitudes. By ‘targeting’ here we don’t mean to imply anything about an outward expression (verbal or otherwise) of the reactive attitudes – indeed, someone could target Sam, in the relevant sense, without Sam’s ever knowing about it. Perhaps you hear about Sam’s misdeed from a friend, and although you and Sam don’t know each other, you feel indignant toward her on the basis of her A-ing. This is sufficient for Sam to be *targeted* by your indignation. It is thus a rather mild form of accountability, but it is accountability nonetheless. Moreover, affirmative answers to the questions about justification and excuse appear to be *sufficient* for an affirmative answer to the question about whether application of the reactive attitudes is in principle justified. Unexempted, unjustified, unexcused wrongdoing entails justified accountability in principle.

It is important not to confuse question 6 with question 7, however:

Question 7: Is any *actual* person *in fact* justified, in the circumstances, in targeting Sam with any of the reactive attitudes?

Question 6 is about justification in the abstract: supposing there is someone in an appropriate position to do so, would it be justified for that person to target Sam with any of the reactive attitudes? Question 7 then asks about the hypothetical supposition: is there someone in an appropriate position to do so? When we ask question 6, we are asking whether targeting Sam with any of the reactive attitudes is in principle

²³ Some theorists understand this distinction in terms of the difference between judging that someone is a morally responsible agent as opposed to judging that it is fair to hold the agent responsible for what he or she has done. For a discussion of how these come apart, see Vargas 2004a, especially pp. 223-226.

²⁴ Note that there are two senses of ‘justification’ at play in our discussion. The first sense appears in question 4, when we are asking whether Sam’s action was morally wrong (whether she had some justification for A-ing). The second sense is at play in the accountability questions about whether anyone is justified in principle in holding Sam responsible.

justified, *even if there is no actual person who is in fact in a position to target her*. To understand better how questions 6 and 7 might come apart, consider the notion of *authority*. In order to *hold someone responsible* for something he or she has done, it is plausible to suppose that the person who is doing the holding needs to possess an appropriate sort of authority. It may be that an accomplice to Sam's morally wrong action, or someone who regularly performs morally wrong actions of the same type, would exhibit hypocrisy were he to target Sam with any particular reactive attitude on the basis of her A-ing. And perhaps potential hypocrisy robs someone of the authority needed for that person to be justified in targeting Sam with any of the reactive attitudes. We can nevertheless ask whether anyone *would be* justified in targeting Sam with any of the reactive attitudes *were he in a position to do so*.²⁵

With questions 1-7 answered affirmatively, we can conclude so far that the following are all true of Sam: 1) She performed A, 2) In performing A, she exercised the capacities that make her a sensible target for both aretaic and reactive appraisal, 3) If someone were in an appropriate position of authority, it would be justified for that person to target Sam with the reactive attitudes on the basis of her A-ing, and 4) Someone in Sam's community is in fact in an appropriate position of authority.²⁶ But none of this says anything about *interaction* between Sam and the members of her moral community. The rest of our questions will take up this issue, as we move to more concrete ways of holding Sam accountable for her action.²⁷

First, however, let us briefly take stock. We have examined the attributability face of moral responsibility, which has both aretaic and

²⁵ It may well be that answers to questions 6 and 7 (and subsequent pairs) will always march in lockstep since it is plausible to suppose that *the agent herself* always occupies the relevant position of authority needed to hold herself accountable for her action. Even if that's right, the distinction between the members of each pair is nevertheless important in second- and third-personal accountability.

²⁶ There may be other factors besides *authority* relevant to whether anyone in Sam's actual moral community is justified in targeting her with the reactive attitudes. We focus on authority here simply because it is one salient factor.

²⁷ Michael Zimmerman (1988) also distinguishes between "inward" and "outward" forms of accountability. The first he calls 'appraisability' and the second 'liability'. For Zimmerman, it seems that appraisability is more a matter of forming the judgment that blame is justified in the circumstances, whereas liability consists in the actual expressions of blame. If we understand Zimmerman correctly, then appraisability in his sense would map onto someone's making the judgment that the answers to questions 6 and following are affirmative. We use the terms 'inward' and 'outward' to refer to two different sorts of *accountability*, whereas Zimmerman's use of those terms marks a distinction between judging blameworthy and actually blaming.

reactive features. We next filled the gap between reactive attributability and accountability by examining the notions of *justification* and *excuse*. Now we are in the middle of examining the accountability face of moral responsibility. The first feature of the accountability face comprises questions about the justifiability, in principle and in practice, of the application of the reactive attitudes. Next, we will continue to examine the accountability face by discussing questions that go beyond mere application of the reactive attitudes.

As with questions 6 and 7, the remaining questions will always come in pairs, with the first member of the pair asking about justification in the abstract, and the second member of the pair asking about justification as it applies to the actual members of Sam's community. We begin with the mildest form of *interaction* that members of Sam's community might have with her in response to her A-ing.

Question 8: Is it justified, in the circumstances, to *outwardly express* any of the reactive attitudes to Sam?

Recall from above that the relevant notion of *targeting* that is at issue in questions 6 and 7 does not necessarily involve any interpersonal interaction. It is sufficient for your targeting Sam with a reactive attitude that you merely resent her for her A-ing, whether or not she knows that you resent her or indeed even knows who you are. The notion of *outward expression* operative in question 8, however, is meant to involve overt interaction. The interaction here may be as simple as your telling Sam that you resent her for her A-ing. This simple assertion would be a way of outwardly expressing your resentment. Alternatively, you might outwardly express your resentment by way of an omission: perhaps you give Sam the cold shoulder.²⁸ Whatever the particular manifestation, though, it is important to realize that outward expression is a stage distinct from mere targeting, because it may raise its own questions about justification. It is one thing inwardly to resent someone; it is quite another to express that resentment. Expressing resentment may raise questions of fairness that are not raised by merely targeting an agent with resentment.²⁹ In addition to the abstract question about justification embodied in question 8, we will also need to ask the following:

²⁸ Of course, in order for the cold shoulder to count as interaction, we must presuppose that Sam and you already know each other and interact on a regular basis. Only in these circumstances would your giving her the cold shoulder have the significance that goes along with outwardly expressing resentment.

²⁹ Watson 1996 contains an excellent discussion of how the notion of *fairness* interacts with the faces of responsibility. See also the interesting discussion in Hieronymi 2004.

Question 9: Is any *actual* person *in fact* justified, in the circumstances, in outwardly expressing any of the reactive attitudes to Sam?

It may be that the authority needed to target Sam with a reactive attitude is not, in itself, sufficient authority to outwardly express any reactive attitude to Sam. If that's right, then a positive answer to question 7 will not automatically generate a positive answer to question 9.

Positive answers to questions 8 and 9 tell us that someone in Sam's moral community is in fact justified in outwardly expressing the reactive attitudes to Sam. But even this is still a far cry from the sort of interaction that constitutes our penal system, so our next questions will move in that direction. Since we are now familiar with the distinction between justification in principle and justification in practice, we will begin to present the questions in pairs:

Question 10: Is it justified, in the circumstances, to impose some sort of sanction (beyond mere outward expression of the reactive attitudes) on Sam on the basis of her A-ing?

Question 11: Is any *actual* person *in fact* justified, in the circumstances, in imposing some sort of sanction (beyond mere outward expression of the reactive attitudes) on Sam on the basis of her A-ing?³⁰

'Sanction' is a slippery term, but we can get the intuitive idea of questions 10 and 11 even without a precise definition by giving some examples. Suppose that the morally wrong action in question is Sam's breaking a promise to you. In accordance with an affirmative answer to question 7, you might resent her but not tell her about it. In accordance with an affirmative answer to question 9, you might tell her that you resent her. But a step further might be for you to *demand* that she compensate you for whatever loss you may have incurred as a result of the broken promise. Or perhaps you might *rebuke* her for her lack of fidelity.³¹ Both of these actions would count as sanctions, as we are using that term in questions 10 and 11. But sanctions might also be

³⁰ For the sake of simplicity and because of the particular case of Sam we are discussing, we phrase these and subsequent questions in terms of negative accountability (blame, sanctions, suffering, etc.), but it should be kept in mind that these stages of responsibility seem to apply to positive accountability as well (praise, rewards, pleasure, etc.).

³¹ We are indebted to Coleen Macnamara for helping us see the significance of interpersonal sanctions such as rebukes and demands.

institutional rather than interpersonal. Imposing a fine or a jail sentence counts as an institutional sanction that may or may not be justified depending on the answers to questions 10 and 11. Again, with respect to sanctions, we need to ask both whether they would be justified if anyone were to occupy a relevant position of authority, and also whether anyone in the moral community in fact does occupy such a position.³²

But even asking about sanctions does not exhaust the possible interesting questions about the sorts of interaction with Sam that might be justified as a response to her morally wrong action. There are at least two more pairs of questions to be discussed.

First, even though we know that someone in her moral community is justified in imposing on Sam some sort of sanction on the basis of her A-ing, the sorts of sanction discussed in connection with questions 10 and 11 do not involve a significant amount of *suffering* on Sam's part. So we might wonder whether measures that are more drastic are justified. This brings us to questions 12 and 13:

Question 12: Is it justified, in the circumstances, to make Sam suffer for her A-ing?

Question 13: Is any *actual* person *in fact* justified, in the circumstances, in making Sam suffer for her A-ing?

Although it is plausible to suppose that even rebukes, demands, fines, and jail sentences should count as actions that cause Sam to *suffer*, in some appropriate sense, there is a more significant sense of suffering that is at issue in questions 12 and 13. Some people, for instance, think that wrongdoers should be made to suffer in precisely the same way and to the same degree as the victim of their wrongdoing. Indeed, support for capital punishment is sometimes fueled by this thought. If Sam's morally wrong action were murder, perhaps killing her would be justified. (We aren't taking a stand on this issue here, of course; we are merely attempting to spell out questions 12 and 13.) The suffering in question need not be as drastic as death, however (which, incidentally, only controversially counts as suffering in any case), and it needn't be "equal" to the wrongdoing. Punishments such as solitary confinement and manual labor may count as the sort of suffering that is significant enough to be relevant to questions 12 and 13. Once again, though, we

³² Since we have moved into the realm of institutional accountability with questions 10 and 11, the term 'person' as it appears in those and subsequent questions should be interpreted broadly enough to include institutions.

need to separate the question of whether such suffering is in principle justifiably imposed on Sam from the question of whether anyone in our moral community is actually in a position to impose it.

Finally, we can wonder about the justifiability of the most extreme sort of punishment: eternal damnation of the sort associated with hell.

Question 14: Is it justified, in the circumstances, to make Sam suffer eternal damnation for her A-ing?

Question 15: Is any *actual* person *in fact* justified, in the circumstances, in making Sam suffer eternal damnation for her A-ing?

Although these questions might seem otiose, there has been some talk in the literature about a sense of responsibility that Galen Strawson has called “heaven-and-hell responsibility”. In an encyclopedia article, Strawson characterizes this sense of (what he calls) “ultimate” responsibility as follows: “‘ultimate’ moral responsibility is responsibility of such a kind that, if we have it, it *makes sense* to propose that it could be just to punish some of us with torment in hell and reward others with bliss in heaven.”³³ We will return to this sense of ultimate responsibility below.

With these final two questions, we have completed our examination of the accountability face of moral responsibility. Before we apply the results of our investigation to debates in the literature, however, it will be useful to codify our results.

4. The Physiognomy of Responsibility

We have so far identified at least 15 distinct analytical features of moral responsibility: 1 precondition, 2 sorts of attributability, 10 questions about accountability, and 2 steps in the space between reactive attributability and accountability.³⁴ Putting all the results together, then, we can construct the following diagram (figure 1). For simplicity, we suppress the distinction between justification in principle and justification in practice as it applies to the accountability face:

³³ Strawson 2004, original emphasis.

³⁴ In this paper, we don’t investigate aretaic appraisal any further than question 2, which is about aretaic attributability. But it is an interesting question whether there are any analytical features on the aretaic side of things that run parallel to what we have called “the space between” and “accountability”. We don’t want to rule out that possibility, but we won’t examine it any further here. Thanks to Randy Clarke for bringing this possibility to our attention.

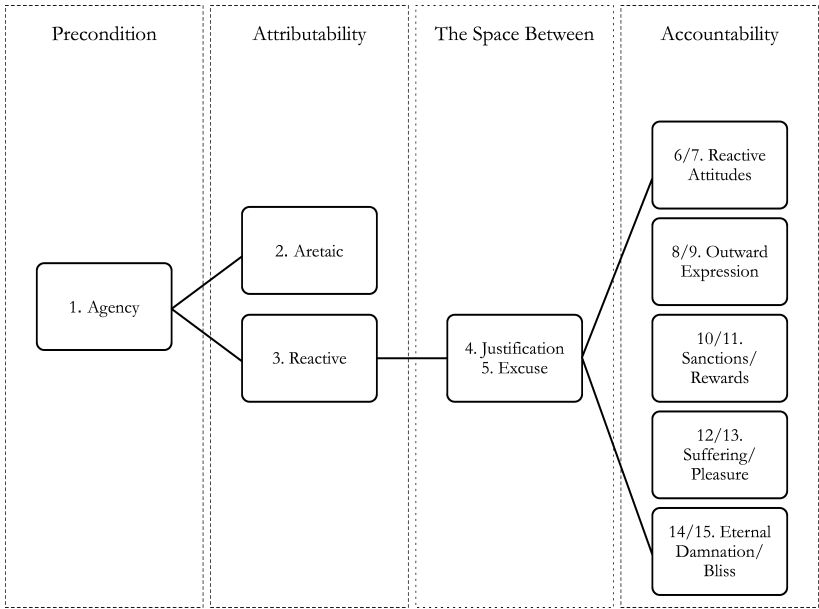


Figure 1: Analytical features of moral responsibility.

Moral responsibility is clearly a rich and multi-faceted concept. Whenever a concept is this nuanced, however, confusion easily arises in discussions of it; this, one might say, is moral responsibility’s frown! In the article from which we have been taking our cue, Gary Watson points out a number of confusions that arose as a result of failing to distinguish the two faces of responsibility.³⁵ We think our detailed physiognomy can help illuminate the literature on moral responsibility in a similar way. We now turn to that task.

5. The Compatibility Question

Prior to Harry Frankfurt’s (1969) pioneering work, which (arguably, at least) shows that there is an important distinction to be made between the ability to do otherwise, on the one hand, and the control required for moral responsibility, on the other, the terms ‘compatibilism’ and ‘incompatibilism’ could be used without qualification. Compatibilists thought that determinism is compatible with both the ability to do

³⁵ For instance, he argues that this failure leads Peter van Inwagen (1983) to conclude incorrectly that skeptics about moral responsibility must contradict themselves, and that the failure also leads Susan Wolf (1990) to posit a suspicious asymmetry with respect to the conditions of praiseworthiness and blameworthiness (the former conditions being less stringent). These are just two of the ways that Watson usefully applies his distinction.

otherwise and moral responsibility; incompatibilists thought otherwise. But those terms are no longer felicitous, since Frankfurt drove a wedge between the ability to do otherwise and moral responsibility. Now theorists must specify what *sort* of compatibilist or incompatibilist they are: whether about determinism and the ability to do otherwise or about determinism and moral responsibility. As a result, someone may be both a compatibilist (in one sense) and an incompatibilist (in another).³⁶

What Frankfurt did for the ability to do otherwise and moral responsibility, Watson did for attributability and accountability. Clarity now demands that compatibilists and incompatibilists about determinism and moral responsibility specify *which face* of moral responsibility is at issue in their theses. Someone may well think that determinism is compatible with attributability even though it is not compatible with accountability.³⁷

In this paper, we are following in Frankfurt and Watson's footsteps by making further distinctions within attributability and accountability. Thus, the terms 'compatibilism' and 'incompatibilism' are slowly becoming more dangerously ambiguous. If our physiognomy is on target, it follows that it is coherent to be a compatibilist about, say, determinism and the sensibleness of aretaic appraisal (question 2) but also an incompatibilist about, say, determinism and the in principle justifiability of eternal damnation (question 14). And there will be many other combinations besides. As a result, clarity demands that theorists be explicit about just what is supposed to be compatible or incompatible with determinism. Lack of precision can lead quickly to confusion and equivocation. Indeed, we turn next to a debate that we think has arisen as a result of failing to distinguish the various analytical stages of moral responsibility properly.

6. Moral Responsibility vs. Blameworthiness

We believe that our apparatus can help diagnose the source of a recent disagreement between John Martin Fischer and Derk Pereboom.³⁸

³⁶ Though this is not the official doctrine of *semicompatibilism*, John Martin Fischer (1994) is a theorist of this sort. Officially, semicompatibilism is agnostic with respect to the compatibility of determinism and the ability to do otherwise.

³⁷ This appears to be the view of Saul Smilansky (2000), though he does not put it in quite these terms.

³⁸ Manuel Vargas reminds us (in personal correspondence) that Fischer is not the only one with whom Pereboom is having this dispute. In particular, Manuel Vargas is also party to the dispute. For Vargas' own thoughts about how to resolve it, see his contribution to Fischer, et al. 2007, especially pp. 212-215.

Pereboom has constructed one of the most influential arguments against compatibilism about determinism and moral responsibility,³⁹ according to which determinism is no less threatening to moral responsibility than external manipulation. There are various “manipulation arguments” in the literature, but Pereboom’s is unique in that it presents a series of four cases that are all supposed to be indistinguishable with respect to the facts of moral responsibility. The first case involves ongoing manipulation by neuroscientists:

Case 1. Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary human being as is possible, given this history. Suppose these neuroscientists ‘locally’ manipulate him to undertake the process of reasoning by which his desires are brought about and modified—directly producing his every state from moment to moment. The neuroscientists manipulate him by, among other things, pushing a series of buttons just before he begins to reason about his situation, thereby causing his reasoning process to be rationally egoistic. Plum is not constrained to act in the sense that he does not act because of an irresistible desire—the neuroscientists do not provide him with an irresistible desire—and he does not think and act contrary to character since he is often manipulated to be rationally egoistic. His effective first-order desire to kill Ms. White conforms to his second-order desires. Plum’s reasoning process exemplifies the various components of moderate reasons-responsiveness. He is receptive to the relevant pattern of reasons, and his reasoning process would have resulted in different choices in some situations in which the egoistic reasons were otherwise. At the same time, he is not exclusively rationally egoistic since he will typically regulate his behavior by moral reasons when the egoistic reasons are relatively weak—weaker than they are in the current situation.⁴⁰

The details of the rest of the cases are similar, except that the second case involves only *initial* (rather than ongoing) manipulation by neuroscientists, the third involves severe influence from the agent’s community, and the last involves mere causal determination. Pereboom argues that since the agent is not morally responsible in the first case, it follows that he is not morally responsible in the last case either, and thus that determinism is incompatible with moral responsibility. It is an ingenious argument, and well worth serious consideration.⁴¹ What’s

³⁹ We beg leave not to follow our own advice about discussing compatibilism without specifying a particular feature of moral responsibility for now, since this is how the disagreement under consideration has proceeded.

⁴⁰ Pereboom 2001, pp. 112-13.

⁴¹ For the details, see Pereboom 2001, especially pp. 110-117.

important for our purposes, however, is the way that John Martin Fischer has responded to the argument.

Fischer maintains that Pereboom's so-called 4-case argument fails because the agent is in fact morally responsible in all four cases, even the first case that involves severe external manipulation. As for the intuition that many people have about the agent's non-responsibility in case 1, Fischer hypothesizes that what we *actually* intuit is the fact that the agent in case 1 is not *blameworthy*, where blameworthiness is distinct from (and has more stringent requirements than) mere moral responsibility. But Pereboom has not been convinced because he thinks the distinction between moral responsibility and blameworthiness is untenable.

To see why Pereboom thinks the distinction is untenable, let's consider how he uses the term 'moral responsibility'. He says that "for an agent to be morally responsible for an action in the sense at issue is for it to belong to him in such a way that he would deserve blame if he understood that it was morally wrong, and he would deserve credit or perhaps praise if he understood that it was morally exemplary."⁴² For Pereboom, blameworthiness is entailed by moral responsibility for an understood morally wrong action. When Fischer admits that the agent in case 1 is morally responsible for a morally wrong act, but contends that he is nevertheless not blameworthy, then, Pereboom hears a contradiction. But Fischer of course does not think that his response to the argument amounts to a contradiction. So what's going on here?

The first step to a diagnosis is to identify just which features of the above physiognomy Fischer and Pereboom intend to pick out by using the terms 'moral responsibility' and 'blameworthiness'. Let's start with Pereboom. We think it is plausible to interpret Pereboom's use of the word 'deserve' along the lines of our notion of justifiability as it applies to the accountability face of responsibility. So, when Pereboom says that an agent deserves blame, it is plausible to suppose that he is saying that it is justified, in principle, to blame the agent. And presumably 'blame' is meant to refer to some particular way of holding the agent accountable for her action. So let's suppose that the word 'blame', in Pereboom's mouth, refers to merely targeting the agent with the reactive attitudes, which is the mildest sort of accountability. To say that an agent is morally responsible for an action, then, in Pereboom's sense, is to say that the action belongs to the agent in such a way that it would be in principle justified to target the agent with the negative reactive attitudes if the

⁴² Pereboom 2006, p. 211.

agent understood that the action was morally wrong or perhaps positive reactive attitudes if the agent understood that the action was morally exemplary. As Pereboom uses the term, ‘moral responsibility’ is a bit about attributability and a bit about accountability. The notion of *belonging* is an attributability notion, whereas the notion of blame is an accountability notion. According to Pereboom, it seems that reactive attributability plus understood moral wrongness (lack of justification and satisfaction of some epistemic condition) entails the justifiability of the mildest form of accountability.

As Fischer uses the term ‘moral responsibility’, an agent can be morally responsible for a morally wrong action and yet not blameworthy for it. It is plausible to interpret this as the claim that a morally wrong action can be attributable to an agent in a reactive sense and yet it may not be justified, even in principle, to target the agent with negative reactive attitudes. We are now in a position to diagnose the source of the disagreement between Fischer and Pereboom.

Whereas Pereboom thinks that positive answers to questions 3 and 4 (together with the satisfaction of some epistemic condition) entail a positive answer to question 6, Fischer thinks that the entailment only goes through if a positive answer to question 5 can be also be had. For the sake of convenience, we repeat the relevant questions here:

Question 3: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of the reactive attitudes?

Question 4: Was Sam’s A-ing morally wrong?

Question 5: Is there no other factor that excuses Sam for A-ing?

Question 6: Is it justified, in the circumstances, to target Sam with any of the reactive attitudes?

In other words, Pereboom thinks that question 5 is not needed. Once you know that a morally wrong action is attributable to an agent in a reactive sense and the agent understood that it was morally wrong (questions 3 and 4 plus an epistemic condition), you are allowed to conclude that the agent can be justifiably held accountable for the action (question 6). Fischer, on the other hand, thinks that there are some interesting cases (Professor Plum’s being one of them) in which

positive answers to questions 3 and 4 (even with an epistemic condition satisfied) do not entail a positive answer to question 6 due to question 5 being answered negatively. Unjustified morally wrong action (understood to be morally wrong by the agent) entails accountability for Pereboom, but for Fischer, only unjustified *and* unexcused morally wrong action (understood to be morally wrong by the agent) entails accountability.

We have provided some evidence above for Fischer's view that question 5 is not superfluous (e.g., the case of the mother who chooses to save her drowning child). Further evidence comes from cases like that of Robert Harris, a severely abused child turned adult murderer. Gary Watson compellingly presents the facts about Robert Harris as a case study in how sensitive our reactive attitudes are to factors like abusive childhoods.⁴³ How ought we to understand Harris's moral responsibility? The answer is of course controversial, but we think it is plausible to give positive answers to questions 3 and 4 and a negative answer to question 5 with respect to Harris. This negative answer to question 5 prevents us from giving a positive answer to question 6. Harris exercised the capacities that made him a sensible target for the reactive attitudes, and he did something that he understood was morally wrong, but (arguably, at least) his formative circumstances get him off the accountability hook (that is, they excuse him).⁴⁴

To make sense of the debate between Fischer and Pereboom, then, we must suppose that Fischer thinks Pereboom's Case 1 is relevantly similar to the cases we have presented as evidence for the necessity of question 5. Because of how Professor Plum is manipulated in case 1, this must prevent us from giving a positive answer to question 5. Professor Plum must be excused for some reason. Is this contention plausible?

⁴³ Watson 1987.

⁴⁴ As Watson points out, this does not mean that we ought to let someone like Harris roam free. Rather, we may still be justified in locking Harris up for consequentialist reasons. Of criminals who are themselves victims, Watson says, "To be sure, [our reluctance to hold them accountable] does not shield victim-criminals from legal sanctions. We still protect ourselves against their murderous assaults; we hunt them down, lock them up, shoot them... Seeing the criminal as himself a victim will not prevent us from shutting the cage or pulling the trigger. But these responses will then tend to seem regulative rather than retributive. In a disconcerting way, they lose their normal expressive function." (Watson 1987, as reprinted in Watson 2004, p. 281). It is worth noting, as an anonymous referee pointed out to us, that it is also plausible to interpret the case of Harris not as a case where accountability has vanished altogether, but instead as a case where blame ought to be *mitigated*. We agree that this is a plausible interpretation of the case and think that an exploration of how formative circumstances might mitigate blame would be valuable, but we will not undertake this large task here. We are merely using the case of Harris to illustrate the particular step of our physiognomy that seems to be causing trouble in the dispute between Fischer and Pereboom.

7. Professor Plum vs. Ernie

To see why it is plausible to suppose that Professor Plum is off the hook for his wrongdoing (and thus is not *blameworthy* though he is *morally responsible*, in Fischer's terminology), it will be useful to compare him with another character from a prominent manipulation argument, Ernie. We will argue that while Plum is plausibly considered off the hook, Ernie is not.

Built into the case of Professor Plum above is the claim that when he kills Ms. White, Professor Plum exercises the capacities that compatibilists (like Fischer) think are required for him to be a sensible target of the reactive attitudes.⁴⁵ According to Pereboom, it is this fact in combination with the fact that Plum has done something he understood to be morally wrong that leads us inexorably to the conclusion that it is justified in principle for Plum to be targeted with the reactive attitudes; in other words, to the conclusion that Plum is blameworthy. But Fischer wants to resist this conclusion because he believes that the way that Plum is manipulated excuses him. Why should Plum's circumstances excuse?

To answer this question, consider the very different sort of manipulation to which our second character, Ernie, is subject. Alfred Mele presents the following manipulation scenario, where Diana is a goddess who oversees a deterministic world:

Diana creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E.⁴⁶

Poor Ernie is the ideally self-controlled agent who develops from Z. Mele uses this case as part of his own manipulation argument to call into question – or pose a significant challenge to – compatibilism (a doctrine about which Mele remains in the end agnostic). We don't, however, need to get into the details here; what's important for our purposes is that compatibilists like Fischer will most likely say that, like Plum, Ernie is morally responsible for A-ing,

⁴⁵ There is some question about whether Professor Plum should even count as an agent *at all*, given the severity of the manipulation to which he is subject. For present purposes, we set aside this worry.

⁴⁶ Mele 2006, p. 188.

incompatibilist protests notwithstanding.⁴⁷ But that's not all Fischer will say about Ernie. Presumably he will also say that Ernie is *blameworthy* for A-ing, as well. (This is one bullet that compatibilists have to bite: the capacities that make us morally responsible agents can be implanted via external manipulation, and sometimes the external manipulation neither justifies nor excuses.)⁴⁸ But now the question is: Why is Ernie blameworthy whereas Plum is not? They are both manipulated into exercising the capacities that make them sensible targets of the reactive attitudes, so how can we explain the difference in blameworthiness?

Despite the fact that both Plum and Ernie are manipulated, the details of the manipulation differ dramatically. Whereas Diana chooses Ernie's "initial conditions" but then leaves them to unfold in the normal way, neuroscientists directly produce Plum's states moment to moment. Plum is thus subject to a much more *invasive* sort of manipulation than is Ernie. Plum is akin to the criminal who had a severely abusive childhood, whereas Ernie is akin to the criminal who had a more or less ordinary childhood. Just as we are inclined to let criminals off the hook to some extent (with respect to our moral practices of blame – *not* necessarily with respect to our institutional practices of incarceration) if we find out that they grew up under tragic formative circumstances, so we should be inclined to let Plum off the hook upon finding out the circumstances under which his desires and values have been created. Or so Fischer would argue.⁴⁹

A full defense of Fischer's response to Pereboom's Case 1 would of course include a detailed specification of just what circumstances do and do not excuse an otherwise morally responsible agent from our accountability practices. But (fortunately!) it is not our aim here to provide a full defense. Rather, we have aimed to use our physiognomy to diagnose the source of the disagreement between Fischer and

⁴⁷ It is striking that, although Mele presents the case of Diane and Ernie as (at least) a challenge for compatibilism, Susan Wolf presents an extremely similar case in which she concludes that the relevant agent, Rose, *is* indeed morally responsible. (See Wolf 1990, pp. 103-116). That is, in Wolf's story, Rose is created by a "Leibnizean-type God" to meet all one's favored compatibilist conditions for freedom and moral responsibility. Without going through the details, it is fascinating – and a bit puzzling – that Wolf thinks it is obvious that Rose is free and responsible, whereas Mele believes that it is quite unclear whether Ernie is free and morally responsible. We are indebted to Justin Coates for this point. Also, there is discussion of Wolf's Leibnizean Story in Fischer and Ravizza 1992 and Fischer 1994, pp. 107-109.

⁴⁸ See Fischer 2008.

⁴⁹ Indeed, Fischer has made claims like this in his 2004.

Pereboom, and to gesture toward a way in which it might be fruitfully resolved. Toward that end, we have suggested that the dispute between Fischer and Pereboom is located at question 5: whereas Pereboom thinks it is not needed, Fischer thinks it is.

It might be worth mentioning an alternative diagnosis of the dispute between Pereboom and Fischer. Recall that by ‘moral responsibility’, Pereboom just means the sort of *belonging* that would make the agent justifiably targeted with negative reactive attitudes on the basis of the action if the agent understood that the action was morally wrong. Above we interpreted this as committing him to the claim that affirmative answers to questions 3 and 4 (as well as satisfaction of an epistemic condition) automatically generate an affirmative answer to question 6. But we could also interpret him as simply using the term ‘moral responsibility’ in such a way that someone is morally responsible for an action *only if* we can give positive answers to 3, 4, and 5. If this is what he means instead, then he and Fischer are not disagreeing about anything; rather, they are talking past each other. On this understanding of ‘moral responsibility’, Fischer will agree that Professor Plum lacks it, but he will insist that Plum still possesses something interesting, namely the capacities that make him a sensible target for the reactive attitudes (recall that this is how Fischer uses the term ‘moral responsibility’). So, it may be that Fischer and Pereboom are disagreeing about question 5, but it also may be that they are simply using the term ‘moral responsibility’ to pick out different features of our physiognomy. Either way, we hope that our diagnosis can help to resolve or, at least, to illuminate their dispute.

8. Heaven-and-Hell Responsibility and Constitutive Luck

Let us now move to another area of the literature where our apparatus may prove illuminating. Recall from above that Galen Strawson has invoked the idea of heaven-and-hell responsibility in characterizing the sort of responsibility that he takes to be at issue in the contemporary debate. For Strawson, we have *ultimate* moral responsibility just in case “it *makes sense* to propose that it could be just to punish some of us with torment in hell and reward others with bliss in heaven.”⁵⁰ Above we speculated that this sort of responsibility is to be identified with accountability step 14, according to which it is justified in principle to make an agent suffer eternal damnation. In fact, however, Strawson may be identifying an analytical step that we did not include in our physiognomy.

Recall that reactive attributability, as we understand it, is the idea that in acting, the agent exercised the capacities that make him or her a

⁵⁰ Strawson 2004.

sensible target for the reactive attitudes (question 3). Another way to put this might be as follows: When an action is attributable to an agent in a reactive sense, it makes sense to propose that it could be just to target the agent with the reactive attitudes. The ‘makes sense’ here captures our notion of ‘sensible target’, and the ‘it could be just’ captures the fact that we only need positive answers to questions 4 and 5 in order to conclude that it is justified in principle to target the agent with the reactive attitudes.⁵¹ Stating reactive attributability in these terms, though, makes it sound parallel to Strawson’s understanding of ultimate moral responsibility. So, perhaps Strawson is introducing yet another sort of attributability: if reactive attributability has to do with whether the agent is a sensible target of the reactive attitudes, then “Strawsonian attributability” will have to do with whether the agent is a sensible candidate for eternal damnation. And just as the capacities required for reactive attributability may differ from those required for aretaic attributability, so they may differ from those required for Strawsonian attributability.

What Strawson’s idea of ultimate moral responsibility might be telling us, then, is that we have missed at least four important sorts of attributability, in addition to aretaic and reactive. Notice that our physiognomy assumes that once we have a positive answer to question 6 (the in principle justifiability of applying the reactive attitudes), no other question of attributability arises as we move up the accountability scale. But perhaps that is wrong. That is, perhaps there will be distinct questions about attributability corresponding to pairs 8/9, 10/11, 12/13, and 14/15, respectively. Presumably they would go as follows:

8/9 Attributability: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of outward expression of the reactive attitudes?

10/11 Attributability: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of sanctions?

12/13 Attributability: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of suffering?

⁵¹ An anonymous referee draws our attention to the fact that whereas Strawson uses the term ‘just’, we have used the term ‘justified’. There is certainly the potential for equivocation here, but we are inclined to think that in this context, both words capture the same core issue, which is (in part) whether responding to a wrongdoer with the reactive attitudes is *fair*.

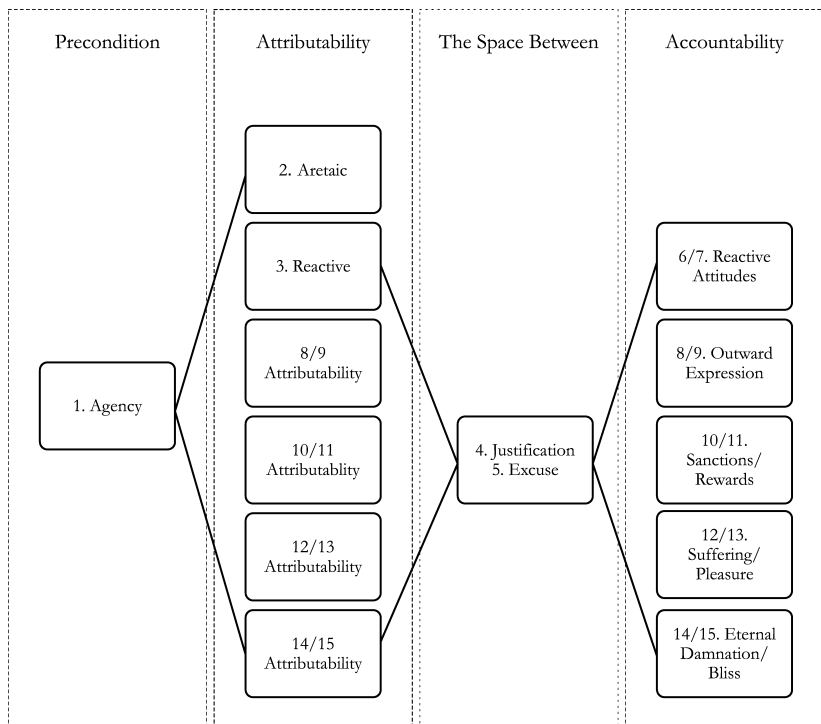


Figure 2: Revised Physiognomy.

14/15 Attributability: In A-ing, did Sam exercise whatever capacities are required for her to count as a *sensible target* of eternal damnation?

If these sorts of attributability really do belong in our physiognomy, then figure 1 would have to be modified so that it looks like figure 2 instead.⁵²

Importantly, it may be that the requirements for 14/15 attributability are more stringent than the requirements for attributability at the lower steps. Indeed, if we look at what Strawson thinks is required for 14/15 attributability, we will see that it is quite a bit more stringent than what is plausibly required for weaker sorts of attributability. We will argue that Strawson's focus on such a high level of attributability has led him to talk at cross-purposes with other theorists.⁵³

⁵² Although we have only included one box for justification and excuse, keep in mind that presumably there would need to be a match between the level of attributability and the level of accountability so that 8/9 attributability wasn't paired with, say, 14/15 accountability. Thanks here to an anonymous referee.

⁵³ Randolph Clarke (2005) has compellingly argued for a similar conclusion about Strawson's arguments. See also Mele 1995, pp. 221-227.

Strawson thinks that whether or not determinism is true, no one is morally responsible for anything because we are all subject to an objectionable amount of *constitutive luck*. To say that someone is subject to constitutive luck is to say, roughly, that the person lacks control over his or her character, personality, temperament, etc. In a word, a person who lacks control over his or her *constitution* is subject to constitutive luck.⁵⁴ The problem with moral responsibility, according to Strawson, is that it requires the absence of constitutive luck. But *everyone* is subject to constitutive luck. So Strawson concludes that no one is morally responsible for anything. One formalized version of his argument runs as follows:

- (1) When you act, you do what you do because of the way you are.
- (2) If you do what you do because of the way you are, then in order to be morally responsible for what you do you must be morally responsible for the way you are.
- (3) But you cannot be morally responsible for the way you are.
- (4) Therefore, you cannot be morally responsible for what you do.⁵⁵

Premise (1) can be given an interpretation according to which it is uncontroversial: we do make our choices, after all, on the basis of what we believe, value, care about, etc. Premise (3) is likewise uncontroversial: we are all subject to constitutive luck because of the way we are intimately situated *within* the world. It is premise (2), therefore, that most opponents of Strawson's argument have attacked. One need not be responsible for the way that one is *in every respect* in order to be morally responsible for what one does. To think that such extreme control is required is to exhibit, as John Martin Fischer has put it, a severe case of "metaphysical megalomania."⁵⁶

But we do not here intend to evaluate the adequacy of Strawson's argument. Rather, we merely want to point out that the use of the term 'morally responsible' in premise (2) is plausibly interpreted, given what Strawson says about moral responsibility elsewhere, as denoting 14/15 attributability. Strawson is here claiming that in order for someone to

⁵⁴ The term comes from Thomas Nagel's excellent discussion of moral luck in chapter 3 of his 1979.

⁵⁵ This way of presenting the argument is adapted from the argument Strawson calls 'Version 1' in his 2002.

⁵⁶ See Fischer 2006, p. 116.

be a sensible target of eternal damnation (or bliss) on the basis of some action, that person must also be a sensible target of eternal damnation (or bliss) with respect to his or her constitution. And since no one's constitution is attributable to them in the sense of 14/15 attributability, it follows (according to Strawson) that no one's actions are attributable to them in that sense either.

Let's suppose, just for fun, that Strawson's argument is sound. Should theorists who believe in the possibility of moral responsibility be worried? They should be *only if* they take themselves to be discussing 14/15 attributability. But virtually no extant theory of moral responsibility even takes a stand on whether our actions are attributable to us in the sense of 14/15 attributability. Again, to take a salient example, consider the theory of John Martin Fischer and Mark Ravizza. Fischer and Ravizza make clear that the sort of moral responsibility they are theorizing about in their book is the sort associated with reactive attributability, or the capacities required for someone to be a sensible target of the reactive attitudes. In principle, then, someone could accept every word of Fischer and Ravizza's theory, and yet think that Strawson's argument for the impossibility of moral responsibility is sound. We speculate that this fact – the fact that he is using 'moral responsibility' to pick out such an extreme notion – is part of why so few theorists are moved by Strawson's argument.⁵⁷

What our discussion of Strawson brings out is that the "conditions for moral responsibility" may differ depending on what particular feature of moral responsibility is under consideration. As we work our way from simple agency (question 1) all the way up to the actual justifiability of eternal damnation or bliss (question 15), the requirements at each level may vary depending on a number of factors, including (1) the agent's capacities, (2) the agent's epistemic situation, (3) the circumstances in which the agent's capacities were developed or exercised, (4) the seriousness of the wrongdoing in question, and (5) the authority possessed by members of the agent's community. (Perhaps there are other factors relevant, too.) We have already seen that the requirements for aretaic and reactive attributability may come apart by focusing on

⁵⁷ Manuel Vargas suggests (in personal correspondence) that we have misunderstood Strawson's reasoning and that what Strawson is really claiming is that reactive attributability requires 14/15 attributability. In this way, Strawson is not talking past other theorists; indeed, he is claiming precisely that other theorists are wrong to think that 14/15 attributability can be neatly divorced from reactive attributability. We admit that this is another plausible way to interpret Strawson, and in this case we would argue that part of why so few theorists are moved by Strawson's argument is that the alleged connection between reactive attributability and 14/15 attributability is left undefended. In fact, Manuel Vargas makes exactly this argument in his 2004b, pp. 414-421.

the case of the psychopath. This would be a difference in factor (1), the agent's capacities. It may be that the requirements for aretaic and reactive attributability also differ with respect to what epistemic situation the agent must be in. Perhaps a consequence of the agent's action needs to be reasonably foreseeable in order for it to be attributable to her in a reactive sense, but not for it to be aretaically attributable to her. This would be a difference in factor (2).⁵⁸

We have also seen that the requirements for reactive attributability for a morally wrong action and the mildest form of accountability for that action may come apart by focusing on the case of Professor Plum. This would be a difference in factor (3), the circumstances in which the agent's capacities were developed. And it is even easier to see how factors (4) and (5) might make a difference. We certainly wouldn't think it justified to make someone suffer for being slightly inconsiderate, though it may be justified to target that person with some degree of resentment. This would be a difference in factor (4). And finally, for a difference in factor (5), consider the fact that it may be that *if God exists*, then he is justified in sending some of us to heaven and some of us to hell. This would amount to the claim that sending us to hell is in principle justified (question 14). But if God doesn't exist, then no one has that authority, thus we would get a negative answer to question 15. And these are just a few ways in which the above five factors might make a difference to what is required for each analytical feature of moral responsibility. The possibilities seem endless. Most importantly, we need to keep these potential differences in mind when in dialogue with other theorists, lest we end up talking past each other.⁵⁹

⁵⁸ In this connection, we suspect that our physiognomy may prove useful in discussions of the epistemic requirements on moral responsibility, as well. It's not altogether clear whether reasonable foreseeability, for example, is a requirement of certain sorts of attributability, though it does seem clearly to be a requirement of certain sorts of accountability.

⁵⁹ Even if it turns out that there is an entailment relationship between certain of the features we have identified (for instance, perhaps the justifiability of targeting an agent with the reactive attitudes entails the justifiability of outward expression of those attitudes), it is still important to separate them for maximum conceptual clarity.

It may be interesting to note that Peter Graham appears to be engaged in a similar project related to the concept of *justification* in epistemology. He distinguishes different types of justification with the aim of preventing theorists from talking at cross-purposes. See both Graham 2007 and Graham 2008. See also Howard Wettstein's *The Magic Prism: An Essay in the Philosophy of Language* (New York: Oxford University Press, 2004), which contains a sustained argument that philosophers who have addressed what they call "Frege's Puzzle" are frequently in fact discussing different puzzles and thus talking past each other.

9. Determinism vs. God's Foreknowledge

Another way our apparatus may prove illuminating is in helping to shed some light on the currently popular thesis that has come to know as *source incompatibilism*. Roughly, source incompatibilism is the view that determinism is incompatible with moral responsibility because determinism is incompatible with *sourcehood*, which is required for moral responsibility. How exactly to spell out this notion of sourcehood is a matter of dispute, but source incompatibilists agree on one thing: sourcehood is *not* to be understood merely as *the ability to do otherwise*. Thus, source incompatibilists are committed to the claim that determinism is threatening to moral responsibility, *quite apart from* the question of whether determinism rules out the ability to do otherwise. But some have had trouble understanding how determinism could be threatening to moral responsibility if not because it rules out the ability to do otherwise.

In reply, source incompatibilists can claim that not all situations in which an agent lacks the ability to do otherwise are relevantly similar. Consider here the difference between determinism and God's foreknowledge. There are plausible arguments to the conclusion that if determinism is true, then no one has the ability to do otherwise.⁶⁰ There are also plausible arguments to the conclusion that if God exists and is everlasting (as opposed to atemporal) and essentially omniscient, then no one has the ability to do otherwise.⁶¹ Suppose that both the arguments with respect to determinism and the arguments with respect to God's foreknowledge are sound. Still, there seems to be a difference between the two cases. Whereas determinism manifests itself as an active part of the causal sequence that unfolds in the world, God's foreknowledge presumably does not (or perhaps, need not) manifest itself in this way. It is open to source incompatibilists, then, to argue that while determinism would rule out moral responsibility *as well as* the ability to do otherwise, God's foreknowledge would not rule out moral responsibility, *despite* its ruling out the ability to do otherwise.⁶²

What's the difference between determinism and God's foreknowledge such that one is incompatible with moral responsibility while the other is not? Here is where our apparatus may be of some assistance.

⁶⁰ See, for instance, Ginet 1966, van Inwagen 1983, and Fischer 1994.

⁶¹ See, for instance, the essays collected in Fischer 1989.

⁶² Similar claims can be made with respect to Frankfurt-examples. Even if we suppose that the presence of the counterfactual intervener does rule out the ability to do otherwise, these cases seem relevantly different from cases in which determinism obtains, since the counterfactual intervener remains idle in the actual sequence in a way that determinism, so to speak, would not. See Fischer 1982 for this point.

The source incompatibilist can make a similar move to the one made by Fischer in response to Pereboom's 4-case argument. Recall that Fischer claimed that although Professor Plum exercised the capacities required for him to be a sensible target of the reactive attitudes when he killed Ms. White, it is nevertheless unjustified to target him with the reactive attitudes because his circumstances excuse him. The source incompatibilist may be able to say something similar about determinism. Although an agent in a deterministic world can perform morally wrong actions that are attributable to him in a reactive sense, the source incompatibilist might say, the fact that he is situated in a deterministic world renders it unjustifiable to target him with any of the reactive attitudes. This amounts to saying that whereas determinism allows for *some* sort of responsibility (namely, reactive attributability), it does not allow for *other* sorts (namely, accountability).

Of course, for the source incompatibilist to make this claim plausible, she will need to give reasons for thinking that determinism creates the sorts of circumstances that get agents off the hook for morally wrong actions. Significantly, though, the source incompatibilist is not stuck with having to claim that determinism would provide an exemption (question 3) or a justification (question 4), a claim that strikes many compatibilists as highly dubious.⁶³ Rather, the source incompatibilist can allow that determinism neither exempts nor justifies, though it does excuse. Again, we don't claim that this way of arguing will work in the end, or even that source incompatibilists will embrace it with open arms, but only that it seems a fruitful avenue to explore given our physiognomy.

10. Compatibilism and the Problem of Evil

We turn now to one last way in which the apparatus of this paper may prove fruitful. Though this is the last way we will consider, we are inclined to think that there are more that would be worth pursuing.

The problem of evil is often thought to be significantly more troubling for compatibilists who believe in free will and moral responsibility than it is for incompatibilists who believe in free will and moral responsibility. To see why, consider this question: Could God have created a world in which free and morally responsible creatures were ensured never to perform morally wrong actions? The backbone of the so-called free will defense against the problem of evil is the claim that the answer to this question is "No". Since God could not have created such a

⁶³ See, for instance, Wallace 1994.

world, in order to allow freedom and moral responsibility, God had to allow for the possibility of wrongdoing. Thus, evil enters the world because God had to make a value judgment, and he chose freedom rather than the ensured absence of wrongdoing.⁶⁴ But compatibilists, it seems, cannot answer the above question negatively. This is because they think that freedom and moral responsibility are compatible with determinism, and if that's right, then God was not forced to choose between a world free from evil and a world with free human beings. He could have created a world with *both* of those features merely by setting the initial conditions of a deterministic world to certain particular values.

One thing that our apparatus helps to make clear, however, is that there are compatibilists and then there are compatibilists.⁶⁵ That is, someone who is a compatibilist about determinism and reactive attributability, for example, may not be a compatibilist about determinism and the in-principle justifiability of retributive punishment. To construct a defense against the problem of evil using our apparatus, then, a compatibilist needs merely to argue that whereas God could have made a deterministic world in which agents perform actions that are attributable to them in a reactive sense, God *couldn't* have made a deterministic world in which agents perform actions for which it is justifiable to make them suffer. And perhaps the sort of capacities that are required for this stronger sort of moral responsibility could not exist in a deterministic world. Therefore, although God didn't have to make a choice between a world with moral responsibility and a world free from evil, perhaps He did have to choose between a world with 14/15 accountability and a world free from evil. And it may well be that the capacities required for 14/15 accountability are so valuable that God chose their actualization instead.

Indeed, given our apparatus, a compatibilist about reactive attributability may well think that the ability to do otherwise is nevertheless required for 14/15 accountability, and thus that determinism would rule out the justifiability of that sort of accountability. (This theorist would thus be an incompatibilist about determinism and 14/15 accountability.) All that is needed for a proper free-will defense, it

⁶⁴ For the classic statement of the free will defense, see Plantinga 1977. For an important expansion upon it, see van Inwagen 2006.

⁶⁵ Of course, as we pointed out above, since Frankfurt's work it has been well known that there are compatibilists (traditional compatibilists) and then there are compatibilists (semicompatibilists). But here what is at issue is not different kinds of freedom (à la Frankfurt), but different kinds – or faces – of responsibility (à la Watson).

seems, is the claim that there is some analytical feature of moral responsibility that is incompatible with determinism. Since almost no contemporary compatibilist is *committed* to being a compatibilist “all the way up”, so to speak, it is in principle open to these theorists to develop a viable version of the free will defense. As before, though, we are not claiming that his project would work, all things considered. But it does seem to be a fruitful project.

11. Conclusion

In this paper, we have closely examined the concept of moral responsibility. Gary Watson has shown that it has two faces, and we have built upon his work by scrutinizing the contours of those faces. We have identified at least some of the rich complexity in the concept of moral responsibility. Although the precise number is somewhat arbitrary, it turns out that (on our analysis) there are at least 15 analytical “stages” or “steps” of moral responsibility: one precondition, two (and possibly six) sorts of attributability, ten stages of accountability, and two features in the space between reactive attributability and accountability. We have gone on to apply the results of our examination to various debates in the literature about moral responsibility.

First, we argued that those who take a stand on the compatibility of determinism and moral responsibility must specify exactly what sort of moral responsibility is under consideration for the sake of clarity. We then argued that the dispute between John Martin Fischer and Derk Pereboom about moral responsibility and blameworthiness can progress by focusing closely on question 5, the question about whether the agent is to be excused from accountability. Next, we looked at Galen Strawson’s notion of ultimate responsibility, which it seems can be better understood as 14/15 attributability. Once we see this, however, we see that Strawson is operating with a distinct sense of ‘moral responsibility’ than most theorists in the debate, a point that makes his argument for the impossibility of moral responsibility significantly less compelling. Finally, we looked at the relationship between moral responsibility and two issues in the philosophy of religion: God’s foreknowledge and the problem of evil. Looking at God’s foreknowledge with our apparatus can help us understand claims made by the source incompatibilist. And finally, by utilizing our apparatus, a certain sort of compatibilist can construct a viable defense against the problem of evil.

In closing, consider the question: Which of the above steps or stages ought we to identify with moral responsibility *simpliciter*? This ques-

tion, we suggest, is misguided. Moral responsibility involves various analytical stages, and once the stage at issue has been specified, there is nothing else that needs to be said. To ask what stage constitutes *real* moral responsibility is to ask an empty question.⁶⁶

We suspect that there are other fruitful ways in which our taxonomy of moral responsibility's facial characteristics might be applied. But the most important point is that there *are* subtle distinctions to be made within the concept of moral responsibility. We suggest that paying close attention to them will improve conceptual clarity for theorists of all stripes.⁶⁷

References

- Arpaly, Nomy. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. New York: Oxford University Press.
- Clarke, Randolph. (2005). "On an Argument for the Impossibility of Moral Responsibility," *Midwest Studies in Philosophy* 29: 13–24.
- Fischer, John Martin. (Ed.). (1989). *God, Foreknowledge, and Freedom*. Stanford, CA: Stanford University Press.
- Fischer, John Martin. (1982). "Responsibility and Control," *Journal of Philosophy* 79: 24–40.
- (2004). "Responsibility and Manipulation," *The Journal of Ethics* 8: 145–177.
- (2006). "The Cards That Are Dealt You," *The Journal of Ethics* 10: 107–129.
- (1994). *The Metaphysics of Free Will: An Essay on Control*. Cambridge, MA: Blackwell.
- (2008). "Review of Alfred Mele's Free Will and Luck," *Mind* 117: 195–201.
- Fischer, John Martin & Ravizza, Mark. (1992). "Responsibility, Freedom, and Reason: Critical Review of *Freedom Within Reason* by Susan Wolf," *Ethics* 102: 368–89.

⁶⁶ We would make similar remarks about the concept of *blame*, as well. Recently, R. Jay Wallace (forthcoming) has criticized T. M. Scanlon's (2008) account of blame because Scanlon does not see the reactive attitudes as an essential component of blame. According to Wallace, Scanlon's account "takes the blame out of blame". But just as it is plausible to suppose that moral responsibility involves various analytical stages, we think it is also plausible to suppose that blame involves various analytical stages, perhaps only some of which will involve the reactive attitudes. To insist that the word 'blame' cannot be used properly in reference to those stages that don't involve the reactive attitudes is perhaps to cleave too tightly to a piece of terminology.

⁶⁷ For helpful comments on previous versions of this paper, we are grateful to Randolph Clarke, Michael McKenna, Angela Smith, Patrick Todd, Manuel Vargas, and the anonymous referees.

- Fischer, John Martn & Ravizza, Mark. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Fischer, John Martin, Kane, Robert, Pereboom, Derk & Vargas, Manuel. (2007). *Four Views on Free Will*. Cambridge, MA: Blackwell.
- Frankfurt, Harry. (1969). "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66: 829–839.
- Ginet, Carl. (1966). "Might We Have No Choice?" In K. Lehrer (Ed.), *Freedom and Determinism* (pp. 87–104). New York: Random House.
- Graham, Peter. (2008). "The Relativist Response to Radical Skepticism," In J. Greco (Ed.), *The Oxford Handbook of Skepticism*. Oxford: Oxford University Press, pp. 392–414.
- (2007). "The Theoretical Diagnosis of Skepticism," *Synthese* 158: 19–39.
- Hieronymi, Pamela. (2004). "The Force and Fairness of Blame," *Philosophical Perspectives* 18: 115–148.
- Irwin, Terence and Fine, Gail. (eds.) (1995). *Aristotle: Selections*. Indianapolis: Hackett.
- Kane, Robert. (1996). *The Significance of Free Will*. Oxford: Oxford University Press.
- Levy, Neil. (2007). "The Responsibility of the Psychopath Revisited," *Philosophy, Psychiatry, and Psychology* 14: 129–138.
- Mele, Alfred. (2006). *Free Will and Luck*. New York: Oxford University Press.
- (1995). *Autonomous Agents*. New York: Oxford University Press.
- Nagel, Thomas. (1979). *Mortal Questions*. New York: Cambridge University Press.
- Pereboom, Derk. (2006). "Reasons-Responsiveness, Alternative Possibilities, and Manipulation Arguments Against Compatibilism: Reflections on John Martin Fischer's *My Way*," *Philosophical Books* 47: 198–212.
- (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.
- Plantinga, Alvin. (1977). *God, Freedom, and Evil*. Wm. B. Eerdmans Publishing Company.
- Scanlon, T. M. (1986). *The Significance of Choice*. The Tanner Lectures on Human Value, delivered at Brasenose College, Oxford University, <http://www.tannerlectures.utah.edu/lectures/documents/scanlon88.pdf>, accessed January 4, 2008.
- (2008). *Moral Dimensions*. Cambridge: Harvard University Press.

- Sher, George. (2006). *In Praise of Blame*. New York: Oxford University Press.
- Smilansky, Saul. (2000). *Free Will and Illusion*. New York: Oxford University Press.
- Smith, Angela. (2007). "On Being Responsible and Holding Responsible," *The Journal of Ethics* 11: 465–484.
- Strawson, Galen. (2004). *Free Will*. (E. Craig, Editor, & London: Routledge) Retrieved November 15, 2007, from Routledge Encyclopedia of Philosophy: <http://www.rep.routledge.com/article/V014SECT4>
- (2002). "The Bounds of Freedom," In R. Kane (ed.), *The Oxford Handbook of Free Will* (pp. 441–460). Oxford: Oxford University Press.
- Strawson, P. F. (1962). "Freedom and Resentment," *Proceedings of the British Academy* 48: 1–25.
- van Inwagen, Peter. (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- (2006). *The Problem of Evil*. Oxford: Oxford University Press.
- Vargas, Manuel. (2004a). "Responsibility and the Aims of Theory: Strawson and Revisionism," *Pacific Philosophical Quarterly* 85: 218–241.
- (2004b). "Libertarianism and Skepticism about Free Will: Some Arguments Against Both," *Philosophical Topics* 32: 403–426.
- Wallace, R. Jay. (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
- (forthcoming). "Dispassionate Opprobrium: On Blame and the Reactive Sentiments," in Rahul Kumar, Samuel Freeman and R. Jay Wallace, eds., *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. New York: Oxford University Press.
- Watson, Gary. (2004). *Agency and Answerability*. Oxford: Oxford University Press.
- (1987). "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," In F. Schoeman (Ed.), *Responsibility, Character, and the Emotions*. New York: Cambridge University Press.
- (1996). "Two Faces of Responsibility," *Philosophical Topics* 24: 227–248.
- Wolf, Susan. (1990). *Freedom Within Reason*. New York: Oxford University Press.
- Zimmerman, Michael. (1988). *An Essay on Moral Responsibility*. Totowa, NJ: Rowman & Littlefield.