

---

# Holding Ourselves Responsible: When What Rightly Matters Doesn't Really Matter



Photo credit: Casey Staff

Lisa Tessman  
BINGHAMTON UNIVERSITY

*Presidential Address delivered at the one hundred twentieth Eastern Division meeting of the American Philosophical Association on January 17, 2024.*

## I. MORAL RESIDUE

For a long time, I have believed that moral failure can be unavoidable. This is a position that many (perhaps most) people reject because it suggests that “ought” does not imply “can”—that we can be morally required to do things that we are unable to do—which, among other things, seems quite unfair. I am not retracting my claim that there are situations in which we can be required to do the impossible. But I want to give due weight to the strong intuition that we can’t be *responsible* for failures to do the impossible. Actually, I want to make room for both the claim that we *can* be responsible for such failures and for the claim that we *cannot* be responsible for them. Unsurprisingly, it is going to be a pluralist position that acknowledges multiple senses of responsibility and kinds of moral failures. I think that the key to understanding these different senses of being responsible is in the asymmetry between what we hold ourselves responsible for and what other people can rightly hold us responsible for, an asymmetry that appears in cases of unavoidable moral failure and in cases in which we hold ourselves responsible for failures to do what other people take to be supererogatory.

Of course, it is standard to think that when people hold themselves responsible even though no one else rightly could, they are making a mistake—perhaps they are being too hard on themselves, or they are experiencing understandable but irrational emotions, but in any case

they are wrong in thinking that they really are responsible for whatever they have done. I reject this interpretation of the phenomenon. I think that we can *really* be responsible simply because it can be fitting to hold ourselves responsible, and this can be true even in situations in which no one else could rightly hold us responsible.

I will call the situations that I am going to examine, and which exhibit this asymmetry, situations of moral residue.<sup>1</sup> Throughout my discussion I will use a single case of moral residue as a sort of touchstone: the case of a physician from Venezuela named Dr. Koana Rojas.<sup>2</sup> Dr. Rojas joined Doctors without Borders to work on a COVID ward early in the COVID-19 pandemic, when no matter what health-care practitioners did, a lot of patients were going to die. In an interview, Dr. Rojas describes her relationship to one particular patient's death. Listening to her words, we can hear how she grapples with her own anguished sense of responsibility for this death, even as others made it clear that they did not hold her responsible:

I had this patient—a 40 year old pediatrician—who had been admitted for Covid-19. From the beginning he had a lot of difficulties breathing. The only thing that could have helped him was to be in intensive care. But we did not have the physical space. It was impossible to transfer him to another hospital because they were all full. There was no other place where he would have received better care. Our last option was palliative care. . . . We saw how his heart rate progressively slowed down. He was not conscious anymore. And he passed away. . . . I remember that his son, while crying and caught in desperation, kept repeating, "I know that you did all you could. Don't worry, doctor. I know you did all you could." But it was hard for me to look him in the eye because deep inside I knew that there were things that I could have done. There were other things we could have tried. We could have transferred him to intensive care, for example. We could have put him on a ventilator. Medically speaking, there were things we could have done. It was not like we did all that was humanly possible. But there are things that go beyond us. The fact that we did not have available beds in the intensive care unit. It is something that does not depend on me.<sup>3</sup>

The deceased patient's son seeks to relieve Dr. Rojas of responsibility precisely because he recognizes that she will tend to hold herself responsible. Dr. Rojas herself goes through several kinds of responses to her role in the patient's death: she probes the question of whether it really was beyond her control by first berating herself for not doing "all that was humanly possible" or "medically" possible, but then acknowledges that these things were not practically possible. She could not have saved the patient, or at least could not have been fairly expected to do "all that was humanly possible" in order to save the patient. But this acknowledgement, although it makes her ambivalent, does not immediately release her from her anguish. She still feels like she has failed.

I take this to be a clear example of the phenomenon of moral residue, marked as it is by the characteristic though peculiar asymmetry between what people hold themselves responsible for and what others hold them responsible for. I define situations of moral residue as situations in which:

- One holds oneself morally responsible for what one thereby takes to be a failure, even though
- other people could not rightly hold one responsible; but nevertheless
- other people might look poorly on one if one does not hold oneself responsible.

In situations of moral residue (as in other situations), it is through having any of several possible emotional attitudes<sup>4</sup> that we hold ourselves responsible. Following Strawson, I will refer to the attitudes through which we hold people responsible as "reactive attitudes," and I will call those of them that we turn toward ourselves "self-reactive attitudes."<sup>5</sup> Strawson identifies "resentment" as a reactive attitude that we direct toward those who have shown us ill will, "indignation" as a reactive attitude that we direct toward those who have shown ill will toward someone other than ourselves (or toward ourselves, if we view ourselves simply as a member of the moral community), and "guilt" as a self-reactive analogue that we direct toward ourselves when we have shown ill will toward someone else. I will use the terms "resentment" and "indignation" to refer to varieties of anger that serve to hold others responsible,<sup>6</sup> and I will speak of guilt, self-blame, or more generally any anguished sense of responsibility as the self-reactive attitudes through which we hold ourselves responsible in cases of moral residue.<sup>7</sup> In the

situation that Dr. Rojas describes, the patient's son's reactive attitude is one that serves to release Dr. Rojas from responsibility, as he expresses neither resentment nor indignation on either his own behalf or on behalf of his father, while Dr. Rojas holds herself responsible with her anguished feeling of having failed the patient.

## II. RESPONSE-DEPENDENCE

The idea of reactive attitudes was introduced by Strawson in the course of his suggesting that being responsible is in some way a function of our practices of holding people responsible, where the practices consist in our reacting to people with certain attitudes. As Gary Watson explains this position: "It is not that we hold people responsible because they are responsible; rather, the idea (our idea) that we are responsible is to be understood by the practice, which itself is not a matter of holding some propositions to be true, but of expressing our concerns and demands about our treatment of one another."<sup>8</sup> Having reactive attitudes, according to Strawson, is a crucial part of our interpersonal relationships, and so it is, he thinks, "practically inconceivable"<sup>9</sup> for us to do without them or without some form of holding others responsible. When we believe that someone is not capable of being a participant in ordinary, adult interpersonal relationships we forbear from having reactive attitudes toward them; we simply take their regard for other people to not matter to us in the relevant respect. But toward all those people who are capable of being responsible to others, and who can thus have the kind of interpersonal relationships that depend crucially on this capacity for responsibility, we have reactive attitudes that reflect the fact that how people like this regard us is something that matters to us; it is a fundamental concern of ours. As Pamela Hieronymi has put it, "being responsible is being such that one's . . . state of mind matters to other people in a certain, distinctive, way."<sup>10</sup> Our reactive attitudes reflect this fundamental concern of ours; treating someone as a responsible agent means treating them as someone from whom we expect certain things.

A theory of responsibility that recognizes that being responsible *is* nothing more than mattering to others in certain ways that are reflected in our responses can be called "response-dependent." Response-dependence rejects response-*independence* about responsibility; response-independent theories claim that facts about who is responsible for what exist in some prior and independent way, and that for a response to be correct it must accurately reflect these pre-existing facts. In a response-dependent theory, in contrast, being responsible depends in

some way on the attitudinal responses through which we hold people responsible. If it were not for the fact that we have fundamental concerns that drive our attitudinal responses to people and to their actions, then people and their actions would not have any normative features related to responsibility: actions would not be right or wrong and nothing could have the feature of being a moral failure for which we are responsible.

This is a sentimentalist claim, because the kinds of responses that can be a source of normativity are emotional, though reason can still play a role in helping us determine which of our emotional responses establish the normative features of the object to which we are responding. As David Hume has suggested in his very compelling example, the action of one person killing another has no normative features—such as being vicious or wrong—that are independent of sentimental responses to the action. In his words: “The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action.”<sup>11</sup> Just like in Hume’s example, in Dr. Rojas’s case, the quality of being responsible—specifically of her being responsible for the patient’s death—is a function of the sentiments, as manifested in the attitudes of the people responding to the situation.

Response-dependent theories can be general rather than pertain only to responsibility, and when they are general they capture the fact that *all* of an object’s normative features—including those that have nothing to do with responsibility—depend in some way on the attitudinal responses that people have to it.<sup>12</sup> For example, according to a broad response-dependent theory, a joke’s being amusing is a function of the fact that people are amused by it; if people were not constituted to have any sense of humor at all, nothing would be funny or amusing. Similarly, an attacker being threatening is a function of the fact that people fear being attacked; if people were not constituted in such a way that they found being attacked to be aversive in any way—if we welcomed or were indifferent to pain or harm or death—then there would be nothing threatening or fearsome in an attacker. Something’s being funny, or being threatening, are functions of people’s attitudinal responses, which reflect the ways in which things matter to them. To return to a response-dependent theory specifically of responsibility, we can say that someone’s being responsible for wronging or failing someone else is a function of the fact that people become resentful or indignant—that is, angry in a certain way—about the kind of thing that we then call a wrongdoing or a moral failure; if we were not constituted to care about how others regarded or treated us or what we did to other people, we

would not experience resentment, indignation, or guilt in reaction, and we would have no notion of responsibility.

Response-dependence, then, is the idea that without creatures who value things, nothing would be valuable; without creatures who experience certain things as required or prohibited, nothing would have any deontic status; without creatures to whom things matter, nothing would matter. Things matter because they matter *to us*.<sup>13</sup> Normative features of objects exist because of how the object matters to people who have emotional, attitudinal responses to the object. More specifically, moral failures are failures because of the way we experience them, given our fundamental concerns.

However, the blatant problem with mere response-dependence, as I have stated it, is that it does not differentiate between attitudinal responses that get it right and attitudinal responses that get it wrong; what I have described is a form of response-dependence that David Shoemaker calls “dispositional response-dependence” in contrast to “normative response-dependence.”<sup>14</sup> According to dispositional response-dependence, whichever reactive attitudes people are ordinarily disposed to have directly establish what we really are responsible for. Dispositional accounts leave us without any way to say that someone’s response is incorrect or *unfitting*, and they do not explain why we in fact do, in our responsibility practices, engage in the normative task of differentiating between fitting and unfitting responses. Just as not everything that people are disposed to laugh at is really funny and not everything that people are typically afraid of is really a threat to them, not everything that people tend to resent or feel guilty about is really a wrongdoing. Sometimes our attitudes get it wrong. Consider a very simple example of this: suppose you walk into the room and sit down without saying hello to me, and, upon observing this, I become hurt and angry at you, and take you to have slighted me. According to any response-dependent theory that lacks a way to differentiate between fitting and unfitting responses, you thus *have* slighted me. But suppose, further, that you did not even see me, and that explains why you failed to say hello; I was mistaken to assume that you saw and ignored me. Although my hurt and angry response suggests that you have slighted me, it would be clear to any alert onlooker that the response is misleading. My attitude has gone wrong, and in this case the reason that it has gone wrong is that it was based on a mistake about a non-normative fact, namely, the fact that you did not see me.<sup>15</sup>

*Normative* response-dependent theories of responsibility, including the version that I am developing here, correct for this problem by

---

differentiating between fitting and unfitting responses and specifying that it is our *fitting* reactive attitudes that establish what people are responsible for. To say that an attitudinal response is fitting is to say that it *correctly* construes the object to which it is a response as having certain normative features; for instance, fitting blame correctly construes the blamed person as having the normative feature of being blameworthy for something. The task for normative response-dependent theories of responsibility is to identify grounds for judging a response fitting or unfitting, that is, grounds for determining whether fittingness conditions have been met. Different versions of normative response-dependence offer competing accounts of what the source of normativity is for determining whether fittingness conditions have been met. Cullin Brown points out that Shoemaker's version of normative response-dependence about responsibility assumes response-dependence at the normative level but not at the metaethical level; Shoemaker, Brown argues, does this because he overlooks the fact that two levels of normative response-dependence are operating in Strawson's position: there is the normative level and the (deeper) metaethical level, which we can think of as the level of normativity itself.<sup>16</sup> The claim of normative response-dependence is essentially a sentimentalist claim, while what Brown calls "metaethical response-dependence" combines normative response-dependence with a response-dependent claim about normativity itself. That is, normative response-dependence makes the sentimentalist claim that the normative features of an object (such as the feature of being a wrongdoing or a moral failure) are a function of what the fitting attitudinal response (such as the attitude of anger or guilt) to the object is. Metaethical response-dependence makes the additional metaethical claim that our responses and their underlying concerns, rather than anything prior to and independent of our responses, are also the sources of the normativity that allows us to determine which reasons of fit are good ones and thus whether fittingness conditions have been met. It is this metaethical point that supports the claim that we are the sources of normativity.<sup>17</sup>

My endorsement of metaethical response-dependence will be central to my account of moral residue.<sup>18</sup> Because metaethical response-dependence locates the source of normativity, namely, the basis for determining whether fittingness conditions have been met, *in us*, for a response to correctly construe an object's normative features the response must correctly express something *about us*—something about *what matters to us*. One might assume that this would leave us without any way to differentiate between fitting and unfitting responses, but it does not. We can differentiate by saying that a correct construal is one that depends on identifying what *really* matters to us, distinguishing

it from what we just mistakenly think is what matters to us. While our responses as a whole reflect what we *take* to matter to us, only a subset of these—our fitting responses—reflect what *really* matters to us. The task will be to identify what it is *in* us or *about* us that allows there to be some distance between all that we experience as mattering and what really matters.<sup>19</sup>

Consider again the simple case in which you walk into a room and do not say hello to me. I *experience* your not saying hello to me as mattering to me in a way that is distinctively reflected in my hurt and angry response. But once I find out the facts—that you did not say hello because you did not see me—I realize that your not saying hello does not really matter to me in the way that I experienced it as mattering to me. Upon realizing this I can recognize that my hurt and angry response was unfitting, and it was unfitting precisely because it reflected what I mistakenly took to matter to me, rather than reflecting what really matters to me.

Cases of moral residue are, of course, much more complicated than this simple case, and my aim is to bring a metaethical response-dependent theory of responsibility to bear on the phenomenon of moral residue in order to determine whether someone really is or is not responsible in situations in which their own self-reactive attitude diverges from other people's reactive attitudes toward them. In Dr. Rojas's case, for instance, the question is whether the patient's death is a failure for which she is responsible. To know that, we need to ask whether Dr. Rojas's response is fitting, and whether the patient's son's response is fitting. If both are fitting, we will have to say that Dr. Rojas both is and is not responsible.

In fact, this is exactly what I am going to say—that she both is and is not responsible—but I realize that this is a hard sell; many people will want to say instead that Dr. Rojas's self-reactive attitude is unfitting—that her emotions have somehow gone wrong and so she is not really responsible. I think that when people reach this conclusion they do so because they are misled by what are called the “wrong kind of reasons” for thinking that it is unfitting for Dr. Rojas to hold herself responsible. Even people who are apprised of the fallacy of mistaking a “wrong kind of reason” for a “right kind of reason” can be misled in this case, because the conflation results from not distinguishing between two different moral considerations that are in play, with each consideration grounding a wrong kind of reason for rejecting the attitude for which the other consideration serves as a right kind of reason. That there are two distinct moral considerations at play goes unnoticed because of a lack of acknowledgment that one of the two moral considerations—

the one that serves as a right kind of reason for Dr. Rojas's self-reactive attitude—exists at all.

### III. THE RIGHT AND WRONG KINDS OF REASONS

The notion of fittingness is built on the distinction between the "right kind of reasons" and the "wrong kind of reasons,"<sup>20</sup> though I find it more illuminating to refer to the "right kind of reasons" as reasons of fit and the "wrong kind of reasons" as "extrinsic reasons."<sup>21</sup> The distinction is drawn because not all reasons in support of an attitude are related in a relevant way to the normative features of the object to which the attitude is a response. If we conflate reasons that bear on the question of what an object's normative features are, which are called the "right kind of reasons" for counting the object as actually having those features, with extrinsic reasons (such as prudential and moral reasons) for having the attitude that one has in response to the object, then we will be unable to infer anything about the normative features of the object just from knowing that the response was supported by good reasons. For instance, I have a reason of fit to fear someone who threatens me with violence; however, I meanwhile have an extrinsic reason to suppress and not feel fear, namely, that fear might cause me to freeze rather than fight or flee, and thus make me more vulnerable. But it is fear (supported by a reason of fit) rather than lack of fear (supported by an extrinsic reason) that reflects the normative features of the attacker. The fact that it is in my interest not to feel fear would be the wrong kind of reason for concluding that the attacker is not fearsome, though it is a perfectly good reason for attempting to suppress my fear.<sup>22</sup>

As Hieronymi has emphasized, a person only takes a consideration to be a reason of fit that bears on a normative question if that consideration matters to them in the relevant way. It matters to me that I avoid being violently attacked, and my fear reflects the way in which it matters: it matters in the specific way that only a fearing kind of attitude construes it as mattering. Our reactive attitudes thus are responses that reflect our normative "take" on the world; settling a normative question (such as "Do you pose a threat to me?" or "Did you wrong me?") on the basis of a reason of fit "amounts to forming the attitude,"<sup>23</sup> so our reactive attitudes are direct reflections of the specific ways in which their objects matter to us.<sup>24</sup> Hieronymi proposes that we distinguish between kinds of reasons "by distinguishing between kinds of questions on which a consideration can bear."<sup>25</sup> A consideration becomes a reason when it stands in a certain relation to the question on which it bears. That is, it becomes a reason because of the relation "between the question on which the

consideration bears and the attitudes of which it counts in favor."<sup>26</sup> Hieronymi suggests that we consider the relation between "*settling a question* and *forming or revising an attitude*."<sup>27</sup> It is when the relation is a constitutive relation that settling the question "amounts to forming the attitude."<sup>28</sup> For instance, my settling the question (in the affirmative) of whether you have wronged me, given the way that your treatment of me matters to me, amounts to my forming the attitude of resentment, so considerations that I take to bear on whether you wronged me in a way that matters to me are what Hieronymi calls constitutive reasons (which I am referring to as reasons of fit) for resenting you. In contrast to developing an attitude on the basis of considerations that serve as constitutive reasons (reasons of fit), when one finds extrinsic reasons for an attitude convincing, finding the reasons convincing is not constitutive of having the attitude. The fact that resenting you will teach you a lesson is an extrinsic reason for resenting you; by being convinced that resenting you will teach you a lesson, I do not thereby form the attitude of resentment, though I might decide to try to cultivate an attitude of resentment. My doing this will not indicate that your treatment of me matters directly to me in the specific way that resentment indicates, though if my attempt to cultivate resentment is successful and I begin to actually resent your treatment of me, that will indicate that it has come to matter to me in this way. This is how acting on an extrinsic reason can lead indirectly—through something coming to matter to me in a new way—to the development of a new constitutive reason (reason of fit).

There are both good and bad reasons of fit, just as there are both good and bad extrinsic reasons (giving rise to the odd-sounding phrases, "bad reasons of the right kind" and "good reasons of the wrong kind"). I will return to the question of what makes a reason of fit a good one, a question that is central because someone's attitude might go wrong by their taking some consideration to be a good reason of fit when it is actually a bad reason of fit, and I will suggest that someone's attitude's going wrong in this way can happen if they experience something as mattering to them in a way that it does not really matter to them. However, I will set this important question aside for now, in order to focus first on the separate problem of conflating reasons of fit and extrinsic reasons, a conflation that might lead us to take some consideration to be a good reason of fit when it is actually not a reason of fit at all. This conflation results in confusion about what the normative features of the object to which we are reacting are, because only reasons of fit can tell us anything about the normative features of the object.

Suppose I hear a joke and feel amusement at it. I have a reason of fit for amusement, namely, my amusement is a direct response to my taking

some consideration as such a reason (whether I am conscious of this or not). If I have a *good* reason of fit for amusement, then the joke is thereby amusing. Recall Hieronymi's point that an attitudinal response such as my amusement is indicative of my settling a question on the basis of a consideration that serves as a reason of fit; settling the question has a constitutive relationship to forming the attitude. When I hear a joke, there is some consideration that bears in this direct and constitutive way, for me, on the question of whether to be amused. Whether it really does bear on amusement in the way that I take it to depends on whether it really does matter to me in the distinctive way that amusement captures (an amusing sort of way), that is, whether I have a *good* reason of fit. Let's suppose for now that I do have a good reason of fit, and thus that the joke has the normative feature of being amusing.

However, one object (such as a joke) might have multiple, distinct, normative features, and the source of each different normative feature of the object is in what matters to me and in each of the distinctive ways in which something about the object matters to me. So far I have just named one response that I have to the joke—amusement—but I might have several distinct responses, each of which might be fitting, because there may be several different considerations about the joke that each matter to me in a distinctive way; each of my distinct emotional responses reflects one distinctive way in which the joke matters to me. Borrowing an example introduced by Justin D'Arms and Daniel Jacobson,<sup>29</sup> let's now suppose that the joke is a racist joke. Because it is morally wrong to support the degradation of people on the basis of race with a positive reaction such as amusement, I have an extrinsic reason—in this case a moral reason—not to be amused. The way in which the joke is morally problematic is an extrinsic reason with respect to amusement, and so it does not bear directly on the question of whether the joke is amusing, just like the fact that becoming fearful when threatened with violence might make me freeze has no bearing on whether an attacker is fearsome.

To think that the moral wrongness of being amused in response to the racist joke would bear on the question of whether the joke is amusing is to commit what D'Arms and Jacobson call the "moralistic fallacy": the fallacy of inferring from the fact that it would be morally wrong to be amused that the joke must not be amusing.<sup>30</sup> Not all conflation of different kinds of reasons are instances of the moralistic fallacy, because the extrinsic reasons might be prudential rather than moral, such as when I have a prudential reason not to feel fear. But when one takes a specifically moral reason to bear on the question of whether

the object also has some different normative feature, one commits the specific error of the moralistic fallacy. D'Arms and Jacobson assume that the different normative feature on which the moral reason is fallaciously taken to bear is always a non-moral feature; I do not share this assumption, and hence I am going to extend their point to cases (including cases of moral residue) in which the reasons of fit for taking an object to have one moral feature are wrongly assumed to bear on the question of whether the object also has some different *but (also) moral* feature.

I am parting ways with D'Arms and Jacobson here because while they are value pluralists, they do not commit to *moral* value pluralism. Their value pluralism applies only to what they call the "sentimental values," which they believe do not include moral values. Sentimental values are shared "anthropocentric values" because of their dependence on a limited range of emotions, namely, those that D'Arms and Jacobson call the "natural emotions" to indicate both that their source is in human nature (which explains why they are pancultural), and that they constitute natural "psychological kinds" rather than social constructions.<sup>31</sup> The plurality of sentimental values, for them, results from the plurality of distinct natural emotions, each of which they believe is at least sometimes fitting.<sup>32</sup> Because I do not link the fittingness of an emotion to the universality of its underlying concerns or to convergence on the judgment that fittingness conditions have been met, I recognize a wider range of normatively (and metaethically) response-dependent values than D'Arms and Jacobson do, many of which I count as moral values.<sup>33</sup> D'Arms and Jacobson accept that one object may have multiple evaluative features, and they note that these different features are constituted by there being different fitting responses to the different features, but they do not consider the possibility that one object may have multiple and incommensurably different *moral* features, which is the key claim of *moral* value pluralism, and the claim that I am making. I maintain that moral features, just like other normative features, are a function of what the fitting attitudinal response is; they are plural because of the plurality of distinctive attitudinal responses that can be fitting.

In the case of the amusing, racist joke, I believe that not only is there a reason of fit for amusement, but there is also a reason of fit for a morally condemning sort of response, namely, a response that construes the joke as morally problematic or as violating a moral norm. If it really does matter to me that people not be degraded on the basis of race and that we not violate norms against racial degradation, then I have a good reason of fit for a morally condemning response. But this means that one of the joke's normative features (its being morally problematic) is

a consideration that can serve as an extrinsic reason *not* to have the attitude that is a fitting response to another of its normative features (its being amusing), and vice versa. The fact that the joke is morally problematic is established by there being a consideration that serves as a good reason of fit for moral condemnation, and the fact that it is morally problematic is a consideration that serves as an extrinsic reason not to be amused by it. Meanwhile, the fact that the joke is amusing is established by there being a consideration that serves as a good reason of fit for amusement, and the fact that it is amusing is a consideration that serves as an extrinsic reason not to morally condemn it, because moral condemnation interferes with our ability to be enjoyably amused (this is the idea that is voiced when people say “just lighten up a bit” or “don’t always be so judgy—it takes all the fun out of everything”).<sup>34</sup>

When there are multiple normative features that conflict or cause ambivalence in the sense that one feature is linked to a response laden with some sort of positive affect and the other is linked to a response laden with some sort of negative affect, the presence of each normative feature is a consideration that functions as an extrinsic reason against having the fitting response to the other normative feature. But it would be a mistake to think that either of the different normative features have any effect on the reasons of *fit* that support the other. And, crucially, it is only reasons of fit that tell us anything about what the normative features of an object are. Thus, if amusement and moral condemnation are each fitting responses even though their valences pull in opposite directions, then the joke really is both amusing and morally wrong. D’Arms and Jacobson dub the response that captures this plurality “fitting ambivalence.”<sup>35</sup>

Now I am in a position to apply this understanding of the problem of conflating reasons of fit and extrinsic reasons to Dr. Rojas’s role in the patient’s death. I believe that Dr. Rojas both is and is not responsible for a moral failure because, as I will argue, both her own self-reactive attitude and the patient’s son’s reactive attitude toward her are fitting. These attitudes are two different, but each fitting, reactions to distinct normative features of Dr. Rojas’s role in the patient’s death; but in this case, unlike in the case of the amusing, racist joke, *both* of the normative features are moral features. There are two different kinds of moral failure for which she might be responsible. Insofar as holding her morally responsible and not holding her morally responsible pull affectively in opposite directions, the presence of one normative feature (the feature of not being morally responsible) gives us an extrinsic reason to discount the reactive attitude toward the other normative feature (the feature of being morally responsible, in a different way), and we might

be tempted to discount this reactive attitude by denying its fittingness. Just like in the case of the amusing, racist joke, each reactive attitude (if it is fitting) establishes a normative feature that in turn becomes a consideration that serves as an extrinsic reason against having the other reactive attitude, but there are distinct reasons of fit in support of each attitude, and we must not conflate extrinsic reasons with reasons of fit or think that the presence of extrinsic reasons against having an attitude somehow undermines the reasons of fit in support of it. Because in this case both normative features are moral features, the insistence that the patient's death both is and is not a moral failure for which Dr. Rojas is responsible requires endorsing not only value pluralism, but moral value pluralism. If both reactive attitudes are fitting, then we will commit a kind of a moralistic fallacy if we think that there being extrinsic reasons for her to not hold herself morally responsible can bear on the reasons of fit for her to hold herself morally responsible. To conceive of this as a version of the moralistic fallacy we must accept (not just value pluralism but) moral value pluralism, and we must keep in mind that the two kinds of moral failures and of responsibility for these failures are two different normative features; they are different from each other—in a way that I will elaborate—just like being amusing and being morally problematic are different normative features of the amusing, racist joke. The discordant reactive attitudes are not just an affirmation and a denial of the same normative feature. And yet a discomfort with ambivalence, perhaps in this case amplified by the fact that the two reactive attitudes seem to be an affirmation and a denial of the very same thing, may yield a tendency to reject the pluralistic claim that a condemning and an exonerating response can be simultaneously fitting.

Calling an attitude fitting depends on the meaningfulness of the distinction between reasons of fit and extrinsic reasons, as well as on there being a meaningful difference between good and bad reasons of fit; a fitting attitude is one that is supported by a reason of fit rather than an extrinsic reason, and one for which the supporting reason of fit is a good reason of fit rather than a bad reason of fit. Thus, defining a fitting attitude as one that correctly construes the object to which it is a response as having certain normative features is equivalent to saying that a fitting attitude is one that is supported by a good reason of fit. And this, in turn, is equivalent to saying that the attitude reflects some particular way in which its object really matters to us.

I will still need to return to the question that I set aside earlier, namely, the question of how to distinguish between a good and bad reason of fit, or, equivalently, how to distinguish between what *really* matters to us and what we might mistakenly take to matter to us. Before

tackling this question, however, I need to say more about what it is to experience something as mattering, and to then say what I think Dr. Rojas experiences as mattering about herself, or more generally, what people might experience as mattering about themselves in cases of moral residue.

I take it that things matter in experience, and that all experience is subjective, so for something to matter, there has to be someone whose subjective experience it matters in. Furthermore, I will make the corresponding metaethical claim: not only is it in our subjective experience that things seem to matter, but it is also in our subjective experience that anything can *really* matter. It is in this sense that fittingness is subjective.<sup>36</sup> “We are the sources of normativity” is not always interpreted subjectively, but when I invoke this phrase I mean to express the idea that it is the subjective experience of what matters that serves as a source of normativity. This is not at odds with taking what is normative to be socially constructed because I take our subjective experiences to themselves be the experiences of creatures who are deeply and complexly social.<sup>37</sup> I will turn next to a discussion of what we have access to from a subjective standpoint, first to understand what we may subjectively experience as mattering, and then, finally, to suggest that the inquiry into what really matters to us, is, ultimately, an inquiry into who we are.

#### **IV. NORMATIVE EXPERIENCE**

I am going to examine two things that people experience as mattering in situations of moral residue, and which are reflected in the two different attitudes which construe one person as, respectively, not responsible and responsible for some moral failure. In the case of Dr. Rojas and her role in her patient’s death, there is one thing that matters to the patient’s son about Dr. Rojas, and there is something different that matters to Dr. Rojas about herself. This is because what tends to matter to us about other people differs from what tends to matter to us about ourselves (and eventually I will suggest that they may both be things that really matter).<sup>38</sup>

Consider the patient’s son first. It matters to him that he hold other people responsible only for that which he can justify holding them responsible for; put differently, the fairness of his normative expectations matters to him. I am thinking of justification roughly the way that some contractualists do: there is a set of normative expectations that are justified in the sense that they would be adopted through some hypothetical, ideally

fair process.<sup>39</sup> Perhaps people imagine what would be “permitted by principles that could not reasonably be rejected” when everyone is similarly motivated to find shared principles.<sup>40</sup> It is tempting to say that it is just *because* they have been arrived at through a fair procedure of justification that normative expectations for what people can rightly be held responsible for determine what people are responsible for. But if we (real, rather than hypothetical) humans are the sources of normativity, then we need to know what matters to us; it is not enough for us to know what would matter to hypothetical, idealized beings, unless it matters to us what *would* matter to them. But in fact, this *does* tend to matter to us, and this—rather than the mere fact that the principles are justifiable under idealized conditions—is the relevant point. I do not take principles or normative expectations to be authoritative just by virtue of being justified. Rather, it is only when (and because) justification *does* matter to us when we hold other people responsible that we have a reason of fit (and potentially a good reason of fit) not to hold others responsible for meeting an unjustifiable normative expectation. That is, if we have internalized fair normative expectations around responsibility, then we will have emotions that reflect this because it will be what actually matters to us.<sup>41</sup> If we had not internalized these normative expectations, then their adoption by idealized beings would not directly bear on how we settle the question of whether someone is morally responsible for violating them, though the fact that other people may interact with us as if these are the operative norms for responsibility would still serve as an extrinsic reason for us, and quite possibly a very good extrinsic reason for us, to apply these norms. I will refer to normative expectations as “shared” only when I mean to indicate that they are expectations that have been internalized by—and thus serve as reasons of fit for—those amongst whom they are shared.

Let’s make the plausible supposition that according to shared normative expectations, we cannot rightly hold others responsible for things that are beyond their control, because to do that would be to say that they morally ought to do something that they cannot do, and a principle requiring this can certainly be reasonably rejected.<sup>42</sup> The patient’s son *does* seem to have internalized this kind of a normative expectation for holding others responsible, in such a way that his genuine emotions and his reactive attitude toward the doctor do reflect what matters to him about her and her actions, and his reactive attitude is in accordance with these shared normative expectations. When he settles the question (in the negative) of whether or not she has transgressed any normative expectation that he has of her, this amounts to his forbearing from forming an attitude of anger or blame. He does not hold her responsible for what was beyond her control.

I am not going to worry about whether the son is making a mistake about what really matters to him when he settles this question, because I assume that most people are not inclined to call the fittingness of the son's reactive attitude into question; it is my claim that Dr. Rojas's attitude may also be fitting that I will eventually need to defend. By assuming that the son's attitude is fitting, we can conclude that Dr. Rojas is not responsible for violating shared normative expectations. Thus we have identified one relevant kind of moral failure as the failure to meet shared normative expectations, and determined that Dr. Rojas is not responsible for having failed in this particular way.

The tougher problem is understanding what Dr. Rojas experiences as mattering to her about her own role in the patient's death. As with the patient's son, I will describe what I think she experiences as mattering to her; but unlike I did with respect to the son, I will not assume that Dr. Rojas has a fitting attitude and will instead return later to the question of whether what she experiences as mattering to her about herself is what *really* matters to her about herself. The fact that shared norms for responsibility (were she to apply them to herself) would not support the judgment that she is responsible is not what seems to most matter to Dr. Rojas (though I will leave open the possibility that this might also matter somewhat to her, thus making her ambivalent). As we hear in her interview, despite the patient's son not holding her responsible, she still cannot look him in the eye because she does not share his exonerating attitude; she feels she has failed in a way that shared normative expectations cannot account for. What, then, does Dr. Rojas's self-reactive attitude reflect if not shared normative expectations about responsibility? I think that it reflects Bernard Williams's insight that "in the story of one's life there is an authority exercised by what one has done, and not merely by what one has intentionally done."<sup>43</sup> This insight into what matters to us about ourselves is central to understanding moral residue.<sup>44</sup>

To get to this understanding, I am going to offer a picture of what I think is a subjective experience of requirement. I call this kind of normative experience a *subjective deontic experience* because it is an experience of being required, and the *what-it-is-like* to have this experience can be known only from a subjective standpoint. To talk about subjective experience, I must at least nod at Thomas Nagel (though he would disagree with much of what I am going to say)<sup>45</sup> because he has given us such a potent account of the uniqueness of the subjective standpoint as the standpoint from which we have experience, contrasting it with an objective standpoint. For Nagel, subjective experience takes place for any creature for whom "there is something that it is like to be

that" creature.<sup>46</sup> In his quite convincing example, we can know all of the scientific facts about bats that are accessible from an objective standpoint, but this does not get us any closer to knowing what it is like to be a bat. Nagel argues that what we subjectively experience is real and irreducible to anything that is accessible from outside of a subjective standpoint, such that a full picture of "reality" must include that which can only be known subjectively. The existence of subjective and objective standpoints entails pluralism and double vision; it raises the question of "how to combine the perspective of a particular person inside the world with an objective view of that same world, the person and his [sic] viewpoint included."<sup>47</sup> I agree with Nagel that "often the pursuit of a highly unified conception of life and the world leads to philosophical mistakes—to false reductions or to the refusal to recognize part of what is real."<sup>48</sup> When we have the sort of subjective deontic experience that I am about to describe, we exercise such double vision, and we can best make sense of it by accepting moral value pluralism. The particular kind of moralistic fallacy that I have identified—namely, the one that consists in inferring from one moral feature of an object that it cannot also have some other distinct moral feature—is a "false reduction" and involves "the refusal to recognize part of what is real." The refusal, in this case, is a refusal to grant that what matters in this subjective deontic experience is something that really matters (as opposed to something that merely seems to matter), and that, assuming metaethical response-dependence, it thus establishes a distinct moral feature of its object.

The deontic experience that I have in mind is among the most powerful experiences of one person mattering subjectively to another, though not all powerful experiences of one person mattering subjectively to another take a deontic form. Rather, I think that when we experience another person as mattering to us in a very compelling way, that experience can take either a more evaluative form or a more deontic form; the more evaluative experiences include the experience of unconditional love and of unbearable loss, and the deontic experiences include the experience of the absolute necessity of protecting someone and the unthinkability of failing to do so.

Although moral residue is best explained as a deontic experience, I am going to take a brief detour first to describe a subjective experience that is evaluative rather than deontic: the experience of the death of someone beloved as being *unbearable* to the griever. I think the two experiences are quite similar because they do both reveal some of the deepest ways in which one person can matter to another and they are, in similar ways, both (apparently) contradicted by how things appear from outside of a subjective standpoint. My hope is that those who have had,

or can imagine having, an experience of the unbearable, and who accept that in subjective experience something *really* can be unbearable, may be able to grant the same reality to the unthinkable.

C. S. Lewis records his experience of his spouse's death as an encounter with the unbearable, writing, soon after her death:

Reality, looked at steadily, is unbearable. And how or why did such a reality blossom (or fester) here and there into the terrible phenomenon called consciousness? Why did it produce things like us who can see it and, seeing it, recoil in loathing?<sup>49</sup>

Authors who write about their experience of unbearable loss do their best to make their own subjective standpoint accessible to others but also may remark on how other people who have not had the experience tend to be unsuccessful at understanding the what-it-is-like to undergo unbearable loss. Joan Didion, for instance, writes:

Grief turns out to be a place none of us know until we reach it. We anticipate (we know) that someone close to us could die, but we do not look beyond the few days or weeks that immediately follow such an imagined death. We misconstrue the nature of even those few days or weeks. We might expect if the death is sudden to feel shock. We do not expect the shock to be oblitative, dislocating to both body and mind. . . . Nor can we know ahead of the fact (and here lies the heart of the difference between grief as we imagine it and grief as it is) the unending absence that follows, the void, the very opposite of meaning, the relentless succession of moments during which we will confront the experience of meaninglessness itself.<sup>50</sup>

Like Didion, I also have this unwelcome knowledge of the what-it-is-like to lose a spouse, though Didion calls it the experience of meaninglessness and in my mind it is the experience of unbearability. But it is, as she says, a matter of facing a relentless succession of moments, with each individual moment utterly unsurvivable.

Because a subjective standpoint is the only place from which unbearability can be a real normative feature of anything, some of these strongest ways in which people matter to us are invisible—in fact, they are denied or dismissed as *impossible*—from outside of a subjective

standpoint. Even the person whose subjective experience it is can step outside of their own subjective standpoint, and see their own experience from the outside—that is, objectively. From an objective standpoint, the subjective experience must be explained away. One might give a causal explanation of why someone in the grip of grief will feel what they feel, but the objective fact is that if something is being borne, then it is not unbearable. I can, with the double vision that Nagel alludes to, occupy my own subjective standpoint and emotionally construe the death of my spouse as having the evaluative feature of being not just bad but *unbearable*, even while I can also shift to an objective standpoint, from which I cannot construe the loss this way, because this construal is, or seems to be, falsified by the non-normative fact that I am bearing the loss; I am, after all, still alive. From the objective standpoint it seems that, given that I am bearing the loss, I must be making a mistake when I interpret my subjective experience as an experience of the unbearable.<sup>51</sup> But just as knowing a collection of facts about bats does not amount to knowing what it is like to be a bat, knowing that the loss is in fact being borne tells one nothing about the what-it-is-like to bear it, the what-it-is-like that makes sense of the seeming paradox of bearing the unbearable.

I think of the experience of the *unthinkable* as a deontic cousin of the evaluative experience of the unbearable: they both reflect the same kind of compelling and deep way in which another person can matter subjectively. From a subjective standpoint, the experience of absolute, even impossible, requirement—of what Harry Frankfurt calls “volitional necessity”<sup>52</sup>—is the experience of any alternative to fulfilling the requirement being unthinkable. When one holds another person’s life in one’s hands, not preserving their life may have the deontic feature of being not just wrong but unthinkable. In evaluative experience, something’s being unbearable is qualitatively different (it has a different what-it-is-like) from its being merely bad; similarly with its deontic cousin: being unthinkable is qualitatively different from being merely prohibited. Just as a death can be unbearable when it is the death of one’s beloved, one’s own failure to prevent someone’s death can be unthinkable for anyone who is in what would normally be a position of responsibility to prevent the death—even if actually preventing it is beyond their control. To echo Williams, it is what one does, and not just what one *intentionally* does, that matters when one does the unthinkable.

One way that we can understand Dr. Rojas’s experience is like this: perhaps it was unthinkable for her to let a patient die when it was humanly possible, even medically possible, to save him. This means that

even if it was *practically* impossible for her to save him because there were no beds available in intensive care—or at least it was impossible without similarly unthinkable sacrifices of the other patients who occupied those beds—she still takes herself to be absolutely required to save him; it simply becomes an impossible requirement.<sup>53</sup> And so she holds herself responsible when he dies. When something is unthinkable we experience it not just as being wrong to *do*, but also as being wrong to even consider doing.<sup>54</sup> If we step outside of Dr. Rojas's subjective standpoint and into an objective standpoint, we observe that Dr. Rojas *does* in fact consider doing, and actually does, the alternative action of placing him in palliative care, so we see that this is, objectively, *thinkable*, just as we see that what is experienced subjectively as unbearable is, in fact, borne.

We can now look at the patient's son's reactive attitude and Dr. Rojas's self-reactive attitude side by side. What matters to the son matters to him because of the way that he has internalized normative expectations about what one may rightly hold others responsible for; this is the kind of thing that tends to matter to us about other people. Assuming that his attitude is fitting, Dr. Rojas really is not responsible for failing in the sense of violating shared normative expectations. What matters to Dr. Rojas about herself is that she not do the unthinkable—that she save the patient, because letting him die when it is medically possible to save him, and humanly possible to save him, even though it is not practically possible to save him, is utterly unthinkable. If her attitude is fitting (which I still need to determine) then she *is* responsible for a different type of moral failure. Namely, she is responsible for having done the unthinkable.

There is one more feature of cases of moral residue that requires explanation before I take up the question of whether a self-reactive attitude like Dr. Rojas's is fitting. In situations of moral residue, even though other people tend not to hold us responsible and could not rightly hold us responsible, it is still true that others might look poorly on us if we do not hold ourselves responsible. This can be explained by the fact that we are capable of taking up other people's subjective standpoints through empathy and imagination, though as Nagel remarks, "the more different from oneself the other experienter is, the less success one can expect with this enterprise."<sup>55</sup> Bats are sufficiently different from humans in what we experience that we cannot successfully take up their subjective standpoint and learn what it is like to be a bat; we are never going to know what it is like to echolocate. But often we *can* know much of what it is like to be another human, and it matters to us that things matter to others, in their subjective experience, in ways that resonate

with how they matter to us. This explains why we expect someone like Dr. Rojas to hold herself responsible even though we do not tend to hold her responsible. We expect her to hold herself responsible because we take up her subjective standpoint in part by imagining what subjective experience we would have if we were her. If Dr. Rojas were to feel no anguished sense of responsibility for the patient's death—if she were to just “shrug it off”<sup>56</sup>—she would essentially be like a bat to us: someone whose subjective reality we cannot grasp. We might judge her harshly for this in a sort of aretaic judgment, seeing her as deficient for not experiencing what we experience.<sup>57</sup> But this is not the same as holding her responsible.

## **V. WHAT REALLY MATTERS (AND WHAT DOESN'T REALLY MATTER)**

So far I have emphasized the contrast between the patient's son's reactive attitude and Dr. Rojas's self-reactive attitude, but, in fact, Dr. Rojas to some degree expresses two conflicting self-reactive attitudes, one of which aligns more with the son's attitude toward her. After reflecting on the fact that she has failed to do all that was “humanly possible,” she also tries to find a place for the thought that “there are things that go beyond us,” such as “the fact that we did not have available beds in the intensive care unit,” characterizing this fact, and perhaps also the patient's death, as “something that does not depend on me.” After all, she too has internalized normative expectations for holding people responsible, and the degree to which she directs these expectations toward herself (rather than only toward other people) may just depend on how swamped she is with her concern with not doing the unthinkable. If Dr. Rojas has two conflicting self-reactive attitudes and both of them are fitting, then she will be in a state of “fitting ambivalence,” with her more condemning attitude and her more exonerating attitude pulling affectively in opposite directions. The mere possibility of her ambivalence being fitting can be obscured through a conflation of reasons of fit and extrinsic reasons, and more specifically by way of the moralistic fallacy, just as the fittingness of our ambivalence toward an amusing, racist joke can be obscured.

Finally, then, it is time to ask what it is fitting for Dr. Rojas to feel. Especially if there is plurality and conflict between different ways that she construes her role in the patient's death, each of which reflects a way in which she experiences something as mattering, how can she know whether one, the other, or both of them are correct? More generally, if we are in a situation of moral residue, how do we determine what *really*

matters to us and separate it from what we might mistakenly take to matter to us?

In speaking of “what *really* matters to us,” part of what I mean to convey is that it is really *us* to whom it matters, and so to know what really matters we will have to ask ourselves who we really are. But then it will become clear that the question of what really matters to us—subjectively—is complicated by the fact that we have no single self or, put differently, no single subjective standpoint from which things matter. We can take up multiple different subjective standpoints that are all in some sense our *own* subjective standpoints. This suggests a path to take in determining what really matters to us: if we think our emotions might be misleading us in our experience of something as mattering, we could consider whether that thing *would* matter to us if only we were different—if only we were perhaps a more ideal self—than the self who experienced that thing as mattering. In other words, maybe that which we experience as mattering to our actual, present self is not what *really* matters to us because there are other selves—selves to whom the same thing would not matter in the same way—that are equally or even more fully who we take ourselves to be.

Consider first that our actual experience might be based on our making mistakes—either about non-normative facts or about how to interpret what we are experiencing—and if eliminating these mistakes would not affect who we are in any meaningful way, then we will want to say that the attitudes that arise as a result of the mistakes are not fitting; they do not reflect what really matters to us. There are simple cases, such as the scenario that I offered in which I made a mistake about a non-normative fact and thought you were ignoring me when really you just did not see me. There can also be cases in which I make a mistake about what my own emotion or attitude is: if you are a few minutes late for our lunch date, I might think I am angry in a way that construes your action as a slight; I will retrospectively recognize that this attitude was misleading if, after eating, I see that I had just been irritable due to low blood sugar (it turns out I was “hangry” rather than angry). In both of these kinds of cases, when we become aware of our mistakes we can identify our own attitudes as unfitting by idealizing ourselves just a bit: we ask what *would* matter to us if we were *ourselves-minus-the-mistakes*, that is, if we were a slightly idealized, only slightly counterfactual, version of *ourselves*. In doing this, we take up the standpoint of the slightly idealized self whom we still recognize as *ourselves* and whose subjective standpoint we can enter.

However, identifying what really matters to us with what would matter to an idealized version of ourself is more problematic when, by idealizing, we idealize away from aspects of ourself that are more meaningful to who we are; we can remain deeply attached to some of these more meaningful even if imperfect aspects of ourself. Thus we may be reluctant to count the standpoint of the self with the counterfactual, idealized traits as still our own subjective standpoint, even if we acknowledge that there are very good extrinsic reasons for trying to change ourselves in the direction of becoming like the self whose attitudes are idealizations. In fact, we only consider these attitudes to be more ideal than our actual attitudes—and we only call them “idealizations”—because we have good extrinsic reasons (which are often prudential reasons) to value having them. But at the same time, the reasons of fit of our non-idealized, actual selves pull in the opposite direction, namely, against developing the attitudes that there are extrinsic reasons in favor of coming to have. The more we identify with or commit to being the self that we actually are, or put differently, the stronger the reasons of fit that we have from the standpoint of our actual, present self, the more reluctant we will be to count as our own the standpoint of an idealized self who no longer had these reasons.

What matters to us is plural and often conflicting. Some of what matters to us drives the idealization and an aspiration to become a different self, and this suggests that what matters from the standpoint of our more idealized self is what we should count as really mattering, so we take the idealized self to really be our self. But at the same time, some of what matters to us drives our resistance to having the kind of transformative experience that would make us into someone to whom the same things would no longer matter in the same way, and this suggests that what matters from the standpoint of our actual self—rather than some transformed self—is what really matters to us, namely, it is what matters to what is really us.<sup>58</sup> This plurality and conflict may leave us both ambivalent and without a clear verdict about what really matters to us and thus on whether to call our actual attitudes fitting; we will not know whether our ambivalent attitudes are fitting, because underneath them is an ambivalence about who we really are—about who to be—and thus about which “matterings” are really ours.

Consider a case in which I have an attitude that I am willing to idealize away from without any sense that this idealized self will be any less me. For example, suppose (counterfactually, as it happens) that I am phobic about snakes, though I rationally believe that garter snakes are not venomous and cannot kill me. I emotionally construe garter snakes as having the evaluative feature of being fearsome. I recognize my own

fear as phobic and might even call my emotions about garter snakes “recalcitrant” in order to indicate that, given the non-normative facts I believe about garter snakes (that they are not venomous and cannot kill me), there is no consideration that serves as a good reason of fit to emotionally construe them as fearsome.<sup>59</sup> I see that there is a good extrinsic reason for me to try to change my fear (which I could presumably do through something along the lines of cognitive behavioral therapy), but I have not yet acted on this extrinsic reason. Thus, when a garter snake crosses my path, I feel fear. Is my fear subjectively fitting? If what really matters subjectively is what *would* matter to the more idealized me—the me who would have brought my emotions into line with my rational belief that garter snakes cannot kill me and thus the me who would respond directly to the snake’s inability to kill me as a good reason of fit for a calm response—then we must conclude (and I think this is plausible) that my fear is subjectively unfitting.

Let’s move to the more complicated example of grief that construes its object, the death of a beloved spouse, as unbearable. Just as in the previous example I had the objective belief that garter snakes cannot kill me, so in this case I can step into an objective standpoint and grasp the fact that I am bearing what I subjectively experience as unbearable, and so my loss must be objectively bearable. The enormous suffering involved in this experience, and the need to function well enough to attend properly to all else that matters in my life (including other people), supply good extrinsic reasons to try to stop grieving in this way.<sup>60</sup> The me who is in some sense more ideal would have acted on these good extrinsic reasons and brought my emotions into line with the objective fact that I am bearing the loss, and so would probably experience a less debilitating kind of grief. So would a milder kind of grief reflect what really matters to me about my loss? After all, we can change our reasons of fit by changing how we feel—by changing how things really matter to us. And even if I have not yet changed, I could, just like in the phobia case, count what would matter to a more idealized me to be what really matters to me. I could take up the subjective standpoint of this more idealized self and count it as my subjective standpoint. If I do so, then we must conclude that my actual response, the response of the non-idealized, actual me, has been unfitting. But unlike in the phobia case, I am inclined to say that neither my response nor Lewis’s nor Didion’s is unfitting. I want to insist that our responses *correctly* construe our spouses’ deaths as having the subjective normative feature of being unbearable.

The difference between fear that construes a garter snake as fearsome and grief that construes a loss as unbearable can be found in the

difference between what would be lost in each case if we were, intentionally or not, transformed into selves without the fear or the grief. If I were phobic about snakes, my fear of snakes would still be no part of who I deeply am, and I would be able to relinquish my double vision about snakes, abandoning the subjective standpoint of the phobic self and replacing it, without significant loss, with the standpoint of *me-minus-the-phobia*. But grief, like love, reveals what matters to me most deeply, but *only* from the standpoint of the actual me. When I—that is, my present self—imagine relinquishing my double vision about the death of my spouse by abandoning my present self’s subjective standpoint, the only standpoint from which the death of my spouse is unbearable, I construe the abandonment of this standpoint as abhorrent because if I did not occupy this standpoint, I would no longer be the self to whom my spouse matters in a way that makes her loss unbearable. Nevertheless, if I maintain my double vision, it also keeps alive the possibility of transforming into a self who no longer experiences unbearability and no longer experiences not experiencing it as abhorrent.

I am far from the first to characterize the experience of grief this way—as something whose diminution one has reason to resist or dread. A passage by Marcel Proust is much cited in what has become a discussion of the “puzzle” of the diminution of grief. Proust writes:

Our dread of a future in which we must forego the sight of faces and the sound of voices which we love and from which today we derive our greatest joy, this dread, far from being dissipated, is intensified, if to the pain of such a privation we feel that there will be added what seems to us now in anticipation more painful still: not to feel it as a pain at all . . . for then our old self would have changed, it would then be not merely the charm of our family, our mistress, our friends that had ceased to environ us, but our affection for them would have been so completely eradicated from our hearts, of which today it is so conspicuous an element, that we should be able to enjoy a life apart from them, the very thought of which today makes us recoil in horror; so that it would be in a real sense the death of the self, a death followed, it is true, by resurrection, but in a different self, to the love of which the elements of the old self that are condemned to die cannot bring themselves to aspire.<sup>61</sup>

The puzzle, as Berislav Marušić describes it, is that the reason of fit for grief—the fact that one has lost the person whom one loves—does not change, and yet the fact that grief diminishes over time seems somehow reasonable. He refers, as have I, to a kind of double vision in order to understand the experience: “The double vision arises from the fact that we cannot apprehend, at once, the object of our emotion together with empirical facts about the emotion, such as the fact, if it is one, that grief is a process.”<sup>62</sup> A similar double vision suffuses situations of moral residue such as Dr. Rojas’s situation. Dr. Rojas cannot apprehend, at once, her responsibility for the patient’s death together with the fact that her anguished sense of responsibility is, according to shared normative expectations, unreasonable. Only double vision can enable her to affirm both.

Thus my sense is that the experience of moral residue will often, though not necessarily always, have more in common with the experience of an unbearable loss than with the experience of a phobia. The unthinkable, like the unbearable, requires a double vision that we cannot will ourselves to be rid of; willing ourselves to be rid of it is abhorrent because to feel an anguished sense of responsibility for having done something that was terrible but was impossible to avoid can, like grief, reveal a compelling and profound way in which one person can matter to another. I suspect that the force and importance of the subjective deontic experience at the core of the phenomenon of moral residue—the sense of requirement that goes beyond what anyone else could rightly require of us—compels us to insist that the actual, present, non-idealized self who has this experience is so deeply who we are that we cannot aspire to transform ourselves into a self who would not have this experience, a self to whom the same things would not matter in the same way. In saying this I am recognizing, as Harry Frankfurt put it, “the importance of what we care about.”<sup>63</sup> We might just not care a lot about the unfairness, according to shared normative expectations, of holding ourselves responsible for what is beyond our control.

### **ACKNOWLEDGMENTS**

This address was delivered at the January 2024 Eastern Division meeting of the American Philosophical Association in New York City.

I thank the European Research Council (ERC) for the Advanced Grant that funds the project on moral residue of which this paper is a part. The project received funding under the European Union’s Horizon Europe Framework Programme (HORIZON) Grant agreement No. 101054147. Views and opinions expressed are, however, those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

I appreciate comments on earlier versions of this paper from colleagues with whom I am working on the project on moral residue: Nikoletta Alexandri, Cullin Brown, Olivia da Costa Fialho, Frank Hakemulder, Albert Christiaan Molewijk, Janine de Snoo-Trimpp, and Jan Helge Solbakk (Principal Investigator). This paper is very much a product of my collaborative work with Cullin Brown, whose thinking has deeply informed my own, and with whom I intend to coauthor a book that will combine the work that I do in this paper with his related work; I thank him especially for his valuable comments on many previous drafts of this paper.

I also am thankful for feedback on the main ideas of this paper from participants at the Boston University Philosophy Department's Ethics Seminar (September 2023); at the Irish Philosophical Society's conference on "Moral Development and Moral Failure" (October 2023); at Colgate University (November 2023); in an informal discussion with my mentees from the Marc Sanders Foundation Mentoring Workshop (2023); and at a talk for the Binghamton University graduate program in Social, Political, Ethical, and Legal Philosophy (December 2023). I also thank my colleagues Randy Friedman and Christopher Morgan-Knapp for particularly thoughtful comments on a draft of the paper, and Oded Na'aman for correspondence on ideas related to the paper.

## NOTES

1. Bernard Williams has famously described the phenomenon that I will be examining, juxtaposing—in the case of the lorry driver who has the bad moral luck of striking and killing a child—two aspects of onlookers' responses to the driver: on the one hand, onlookers will feel an impulse to assure the driver that he is not responsible; on the other hand, onlookers normatively expect that the driver will hold himself responsible, as anyone with adequate—and ordinary—moral sensibilities will tend to do:

The lorry driver who, through no fault of his, runs over a child, will feel differently from any spectator. . . . Doubtless, and rightly, people will try, in comforting him, to move the driver from this state of feeling, move him indeed from where he is to something more like the place of a spectator, but it is important that this is seen as something that should need to be done, and indeed some doubt would be felt about a driver who too blandly or readily moved to that position. (Williams, *Moral Luck*, 28)

Justin D'Arms and Daniel Jacobson argue that it is unfitting for the lorry driver to feel guilt (*Rational Sentimentalism*, 171–75) and that "an especially astute philosopher [i.e., Williams] has been led into error, or at least imprecision, because he has mistaken a good reason of the wrong kind for a reason of fit" (*Rational Sentimentalism*, 175). In other words, it is because of this error that Williams takes the lorry driver's response to be fitting. Oddly, my argument will be precisely the opposite. As a preview: I will argue that those who *don't* recognize the fittingness of the lorry driver's anguished sense of responsibility may do so because they commit the version of this error that D'Arms and Jacobson name the "moralistic fallacy"; they infer from the fact that it would be morally wrong for anyone else to hold the driver responsible that his own sense of responsibility must be unfitting (though this is not why D'Arms and Jacobson themselves find the lorry driver's guilt to be unfitting). As will become clear, however, my argument will require several claims that D'Arms and Jacobson would reject.

2. In the health-care field, and especially in literature on nursing ethics, a closely related phenomenon is often called "moral distress." However, because definitions of moral distress vary and because most of the literature on moral distress in health care aims at empirically studying the distress for the purpose of preventing or treating it—and my aims are different—I will continue to use the

term “moral residue” rather than “moral distress” to refer to the phenomenon, even in the context of health care.

3. Excerpt from an interview with Dr. Koana Rojas, taken from the English subtitles on the interview video, which is available from Médecins Sans Frontières / Doctors without Borders, <https://www.doctorswithoutborders.org/latest/silent-wounds-exploring-moral-and-ethical-challenges-covid-19>.
4. Hereafter I will simply refer to them as “attitudes,” but my claims apply only to attitudes that are emotional (and not, for instance, to beliefs about non-normative facts), and I reject a cognitivist account of what emotions are. A cognitivist account, as D’Arms and Jacobson summarize it, claims that “*emotion types are constituted by emotion-independent thoughts necessary for having the emotion, which differentiate them by identifying their evaluative content*” (*Rational Sentimentalism*, 90; italics in the original). The response-dependent theory of responsibility that I adopt is sentimentalist and requires a rejection of a cognitivist theory of emotions. However, I do not depend, as D’Arms and Jacobson (and some other sentimentalists) do, on the claim that the emotions that are relevant as sources of normativity are limited to those that are (near-) universal (and not just pancultural) due to their being part of human nature (see D’Arms and Jacobson, *Rational Sentimentalism*, 116–25).
5. Strawson, “Freedom and Resentment.”
6. D’Arms and Jacobson argue that a subclass of emotions, including resentment, can be characterized in terms of their cognitive content or judgments; they refer to these as “cognitive sharpenings” of an emotion (D’Arms and Jacobson, “The Significance of Recalcitrant Emotion (or, Anti-Quasijudgementalism),” 137–38). Resentment, they suggest, includes a belief about being wronged (“The Significance of Recalcitrant Emotion (or, Anti-Quasijudgementalism),” 143). If resentment is cognitively sharpened in this way, then using the term would support response-independence, because to say that someone had such an emotion would be to describe them as having a response to the prior fact of someone being responsible for a wrongdoing. I am not using the term “resentment” in this way; namely, in keeping with how Strawson uses the term, I do not mean to indicate that the person who feels resentment must have a prior belief about the object to which they are responding being a wrongdoing. I am using the term simply to denote the variety of anger that serves to hold others responsible, roughly what David Shoemaker calls “agential anger” (*Responsibility from the Margins*, 90) or “blaming anger” (“Response-Dependent Theories of Responsibility,” 314) to differentiate it from the anger that one feels as a reaction to frustration. However, unlike Strawson, I do not distinguish between resentment and indignation by taking resentment to be a personal *rather* than moral reaction, and indignation to be a moral reaction (either on one’s own behalf or on behalf of others) (see Strawson, “Freedom and Resentment,” 200).
7. Other terms have been used to characterize the emotions people experience in situations of moral residue. Williams (*Moral Luck*) introduces the term “agent-regret” to refer to the sense of moral responsibility someone has when (the outcome of) their own action was at least partly due to luck. Stephen De Wijze (“Tragic-Remorse—The Anguish of Dirty Hands”) proposes the term “tragic-remorse” to characterize the emotion indicative of moral responsibility in cases of unavoidable moral wrongdoing. Each of these terms is too narrow to cover the range of situations I have in mind. The self-reactive attitudes in situations of moral residue vary. For instance, one might hold oneself responsible in a self-blaming way, or alternatively one might hold oneself responsible in a way that is compatible with not taking oneself to be blameworthy. These different responses construe their objects somewhat differently, and if they are each fitting responses to their respective objects, then they establish that their objects (namely, oneself as an agent, along with one’s actions) really do have different normative features. In the first case, one has an attitude that construes oneself as

blameworthy, and if this response is fitting, then one really is blameworthy, and in the second case, one's attitude construes oneself as having failed in some non-blameworthy way, and if this response is fitting, then one really is responsible in some non-blameworthy way for one's failure. Different cases of moral residue can be classified into different *types* of moral residue cases according to these different normative features, where these features have in turn been established by determining how a fitting response construes the agent. There are important and interesting differences between types of moral residue cases, but I will not be exploring these differences here; however, this address is a part of a larger interdisciplinary project <https://www.med.uio.no/helsam/english/research/projects/more/index.html> in which a team of researchers, led by Jan Helge Solbakk (<https://www.med.uio.no/helsam/english/people/aca/janhs/>), are (empirically and philosophically) investigating these sorts of differences amongst health-care workers' experiences of moral residue.

8. Watson, *Agency and Answerability: Selected Essays*, 222.
9. Strawson, "Freedom and Resentment," 197.
10. Hieronymi, "Introduction" to *Minds that Matter*, 10.
11. Hume, *Treatise*, 3.1.1.
12. When I refer to "normative features" of an object, I am using the term "normative" to cover both evaluative and deontic features, where evaluative features are a function of a particular way in which something is valued or disvalued (or taken to be good or bad) and deontic features are a function of a particular way in which something is taken to be required or prohibited. But anyone who objects to my use of "normative" may simply substitute "evaluative and/or deontic features" for "normative features."
13. Harry Frankfurt puts it directly: "We are creatures to whom things matter" (*The Importance of What We Care About*, 80). This conception of normativity in terms of mattering and specifically in terms of what matters to us (as the sources of normativity) is also present in the work of Humean constructivists such as James Lenman and Sharon Street (as well as myself in Tessman, *Moral Failure: On the Impossible Demands of Morality and When Doing the Right Thing Is Impossible*). As Lenman puts it: "We are creatures who care about stuff" ("Naturalism without Tears," 1). For Street, normative experience is "the experience of various things in the world as 'counting in favor of' or 'calling for' or 'demanding' certain responses on our part" ("Constructivism about Reasons," 240). She characterizes us—beings who have such experiences—as the sources of normativity, saying, for instance, that value "is something conferred upon the world by valuing creatures, and it enters and exits the world with them" ("Coming to Terms with Contingency: Humean Constructivism about Practical Reason," 40). The conception of normativity in terms of what matters to us has also been introduced into work on responsibility in the Strawsonian tradition by Hieronymi, who claims that the fittingness of our responses can be better understood in terms of *mattering* than in terms of *meriting*; as Hieronymi puts it, "mattering, not meriting, is . . . central" to moral responsibility ("I'll Bet You Think This Blame is About You," 60).
14. Shoemaker, "Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame"; "Response-Dependent Theories of Responsibility."
15. Dispositional response-dependent theories might avoid objections based on this sort of example by specifying that being responsible depends only on how we are disposed to hold people responsible *under certain conditions*—perhaps those that do not lend themselves to mistakes about non-normative facts. Shoemaker builds this into the thesis of dispositional response-dependence about responsibility: "X is blameworthy (and thus responsible) for some action or attitude A if and only if, and in virtue of the fact that, people are disposed to respond to X with blaming anger for A in certain standard conditions"

("Response-Dependent Theories of Responsibility," 315). However, Shoemaker then shows how untenable this specification is, noting that "it is entirely obscure what the 'standard conditions' could be" ("Response-Dependent Theories of Responsibility," 315). One cannot unpack "standard conditions" without importing normativity, thus turning the theory into a normative rather than solely dispositional theory.

16. Brown, "Two Levels of Response-Dependence about Responsibility"; Shoemaker "Response-Dependent Responsibility"; "Response-Dependent Theories of Responsibility."
17. A bit more detail about the differences between Shoemaker's and Brown's positions: Shoemaker's normative response-dependent theory neither accepts nor rejects what Brown calls "metaethical response-dependence" but instead remains neutral about the source of the normativity that allows us to determine whether fittingness conditions have been met. Shoemaker defines normative response-dependence, whether about responsibility or any other response-dependent normative feature, with reference to *merit*, because he equates fittingness with merit, which is itself a normative concept. Focusing on blameworthiness as a form of responsibility and on a certain kind of anger as a blaming response, Shoemaker claims: "X is blameworthy for some action or attitude A if and only if, and in virtue of the fact that, X merits blaming anger for A" ("Response-Dependent Theories of Responsibility," 318). One then must ask what makes a reactive attitude fitting or merited. Shoemaker argues that "slighting someone" is "the fitting anger-maker," and that "what makes slighting someone . . . the fitting anger-maker is, constitutively, our sensibilities" in the sense that slighting someone is the sort of property "to which refined human anger sensibilities tend to respond with blaming anger" ("Response-Dependent Theories of Responsibility," 319). In considering how or why our refined sensibilities provide the kind of normativity needed to establish that some relation is one of merit or fit, however, Shoemaker leaves open several possibilities, arguing that although "the relevant normativity may be found in the *refined* sensibility . . . a response-dependent theorist can be neutral about what determines refinement in a sensibility" ("Response-Dependent Theories of Responsibility," 319). For Shoemaker, then, a normative response-dependent theorist can be neutral about the source of normativity; the normativity *itself* need not have as its ultimate source our emotional or attitudinal responses. For instance, Shoemaker suggests, such a theorist could make the teleological claim that there is "a distinctly human good, generating reasons independent of our attitudes and particular sensibilities to develop ourselves in order to promote or achieve it" and argue that our refined sensibilities are those that underlie the attitudinal responses that contribute to this human good ("Response-Dependent Theories of Responsibility," 319). For such a theorist, "the story about refinement would be response-independent" ("Response-Dependent Theories of Responsibility," 319) and the theory would be response-dependent only in a limited sense: specific normative features of an object, such as the feature of being responsible, would still depend on attitudinal responses, but the normativity itself comes from the response-independent, teleological notion of the human good; as Shoemaker puts it, "its normativity would transfer" ("Response-Dependent Theories of Responsibility," 319–20) to the response-dependent account of how, for instance, our responsibility-responses can fittingly serve to hold people responsible. For Shoemaker, then, a theory of responsibility can be a normative response-dependent theory but still be, ultimately, dependent for its normativity on a response-independent account of which sensibilities are refined and thus which responses are fitting.

Brown both endorses metaethical response-dependence and argues that in order to defend the position that he interprets Strawson to take, we need to commit to metaethical response-dependence. Brown points out that the challenge of defending a Strawsonian account of responsibility against response-independent views requires this thoroughly (i.e., metaethically) response-dependent view:

Strawson's emphasis on our actual, ordinary reactive attitudes suggests that what *actually matters to us* plays an important role at the level of normativity—that what actually matters to us bears on whether a reactive attitude's fittingness conditions are *met*—and this is incompatible with a response-independent conception of normativity. Now, Strawsonians must tread carefully here, as the claim cannot be that our actual, ordinary reactive attitudes simply *constitute* the realities of moral responsibility; Strawsonians need a way to cast out some such attitudes and vindicate others if the view is to have any normative bite. The claim, then, must be that the standards for assessing our reactive attitudes ultimately come from *us*—from what *really matters to us* and from the various ways in which our reactive attitudes can reflect this or fail to do so. (Brown, "Two Levels of Response-Dependence about Responsibility," 25)

Furthermore, Brown argues that without metaethical response-dependence, Strawson would be unable to show that the truth of determinism could never undermine the possibility of responsibility; for instance, if someone who has a refined sensibility (as judged by response-independent criteria) "would never experience blaming anger, given the truth of determinism" (Brown, "Two Levels of Response-Dependence about Responsibility," 29), then, contrary to what Strawson aims to show, determinism would be incompatible with responsibility.

18. In Tessman, *Moral Failure*, I develop a Humean form of metaethical constructivism that shares with metaethical response-dependence the core idea that we are the sources of normativity; normativity is a function of what really matters to us.
19. Strawson's own theory, while vague about the details, offers both an account of what we experience as mattering to us, and an acknowledgement that there is a "critical gap" between what we experience as mattering—as reflected in our actual attitudes—and what we might want to say really matters to us, which he believes can be identified through a process of criticism and justification that is internal to our practices of responsibility rather than response-independent. He remarks: "Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external 'rational' justification" ("Freedom and Resentment," 208). I take myself to be participating in this internal process of criticism, in that I will be making the critical point that when Strawson observes what matters to us he overlooks something that matters to us about ourselves and our own actions (as reflected in our self-reactive attitudes) that does not matter to us in the same way about other people (as reflected in our reactive attitudes towards other people).
20. Robinowicz and Rønnow-Rasmussen, "The Strike of the Demon: On Fitting Pro-attitudes and Value."
21. Hieronymi, "The Wrong Kind of Reason."
22. I have altered D'Arms and Jacobson's example of the wolf: "You mustn't let it sense that you're afraid' is a good reason not to fear the approaching wolf; but, unfortunately, it is no reason to doubt that the wolf is fearsome" ("The Moralistic Fallacy: On the 'Appropriateness' of Emotions," 87).
23. Hieronymi, "The Wrong Kind of Reason," 447.
24. Though they have disagreements with each other, the idea that considerations can serve as reasons of fit only when they provoke direct and in some sense nonvoluntary emotional responses is present in the work of both Hieronymi and D'Arms and Jacobson. See Hieronymi ("The Wrong Kind of Reason"; "I'll

Bet You Think This Blame is About You”) and D’Arms and Jacobson (*Rational Sentimentalism*, chapter 4).

25. Hieronymi, “The Wrong Kind of Reason,” 438.
26. Hieronymi, “The Wrong Kind of Reason,” 438.
27. Hieronymi, “The Wrong Kind of Reason,” 447; emphasis in the original.
28. Hieronymi, “The Wrong Kind of Reason,” 447.
29. D’Arms and Jacobson, “The Moralistic Fallacy.”
30. D’Arms and Jacobson, “The Moralistic Fallacy”; *Rational Sentimentalism*.
31. D’Arms and Jacobson, *Rational Sentimentalism*, 106–25. They offer a “provisional list” of natural emotions: “amusement, anger, contempt, disgust, envy, fear, guilt, jealousy, joy, pity, pride, regret, shame, and sorrow” (D’Arms and Jacobson, *Rational Sentimentalism*, 107). They note that anger and guilt can both be fitting responses to non-moral objects, so their inclusion of these emotions on the list does not undermine their claim that moral values are not sentimental values (D’Arms and Jacobson, *Rational Sentimentalism*, 185; 203–12). However, they classify actions such as personal betrayals (to which anger can be a fitting response) as nonmoral; I employ a wider conception of morality than they do, and so would allow for personal betrayals to count as moral failures (see my note #33).
32. D’Arms and Jacobson argue that the “deep and wide” concerns that are reflected in the natural emotions provide anthropocentric reasons of fit for these emotions (*Rational Sentimentalism*, 50), though deep and wide concerns provide good reasons of fit only if these reasons are not “undermined by critical reflection” (*Rational Sentimentalism*, 55). They believe that under good conditions, humans would (hypothetically) converge on what they recognize as good reasons of fit for these emotions, and this is what warrants calling them “anthropocentric reasons” (*Rational Sentimentalism*, 79–86).
33. While I appreciate the ways that D’Arms and Jacobson insist that the narrowly moral does not exhaust all that is valuable for humans, I think that the contrast between moral (narrowly understood) and nonmoral is misleading, because some of what we want morality to do is done best by the sorts of emotional and attitudinal responses that are outside of the narrowly moral, and so it makes more sense to describe morality itself as pluralistic and as covering a broad range. I thus use the label “moral” for all values and requirements that contribute to what David Wong refers to as the functions of morality; for Wong, one of the functions of morality is “to regulate cooperation, conflicts of interest, and the division of labor and to specify the conditions under which some people have authority over others with respect to cooperative activities” (*Natural Moralities: A Defense of Pluralistic Relativism*, 37), and morality also “serves the function of promoting a psychological order within the individual, and not just between individuals who are cooperating with each other” (*Natural Moralities*, 40). We will tend to experience values or requirements that serve these functions as having what Margaret Urban Walker calls “the specifically *moral authority of morality*” (“Seeing Power in Morality: A Proposal for Feminist Naturalism in Ethics,” 8; italics in the original). I develop this point in Tessman (*Moral Failure; When Doing the Right Thing Is Impossible*) as part of a Humean metaethical constructivist theory.
34. Susan Wolf has made a similar point about the way that moral values tend to “crowd out” other values, noting, “A moral saint will have to be very, very nice. It is important that he not be offensive. The worry is that, as a result, he will have to be dull-witted or humorless or bland” (*The Variety of Values*, 14).
35. D’Arms and Jacobson, *Rational Sentimentalism*, 164.

36. D'Arms and Jacobson must deny this, at least as applied to the "anthropocentric values," because for the natural emotions that establish these values, they claim that fittingness is determined by the (hypothetical) convergence on judgments about when fittingness conditions have been met (see my note #32). While I agree that we are likely to observe a high degree of convergence in what really matters subjectively to different people (under the right conditions), such convergence is not a condition of the fittingness of their emotions or attitudes.
37. This claim is developed in Brown, "Subjectivism for Deeply Social Beings."
38. Strawson overlooks the fact that we tend to hold ourselves responsible for actions or outcomes for which we do not tend to hold others responsible, arguing that we hold ourselves responsible, through self-reactive attitudes, in a way that mirrors how we hold others responsible ("Freedom and Resentment," 200–201). The reactive attitudes through which we hold people responsible, according to Strawson, reflect just one thing that matters to us about people: "quality of will," namely, "attitudes towards us of goodwill, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other" ("Freedom and Resentment," 191). Our reactive attitudes include resentment towards those who offend against us and its "vicarious analogue," indignation towards those who offend against either ourselves or others. Strawson also divides our attitudes towards other people into the "personal" and the "moral" reactive attitudes, and one's reactive attitudes on one's own behalf might be either personal—including such things as hurt feelings in the case of unrequited love, where there is no moral reason to expect to be loved—or moral—such as indignation indicating that one has been morally wronged. Strawson pays attention to the fact that we sometimes hold ourselves responsible but argues that we hold ourselves responsible in a way that is an analogue of how we hold others responsible; just as indignation is the "vicarious analogue" of resentment, attitudes such as "a sense of obligation" and "guilt" are self-directed analogues ("Freedom and Resentment," 200–201). They occur when "we demand . . . of ourselves for others, something of the regard which we demand of others for ourselves" ("Freedom and Resentment," 201). Other Strawsonians develop this line of thought, arguing that we hold ourselves responsible by adopting the position of any member of the moral community even toward ourselves; we adopt what Stephen Darwall calls the "second person standpoint," namely, "the perspective you and I take up when we make and acknowledge claims on one another's conduct and will" (*The Second-Person Standpoint: Morality, Respect, and Accountability*, 3). Darwall conceives of the relevant form of holding oneself responsible as "second-personal competence," which he defines as "the capacity to make demands of oneself from a second-person standpoint," and further explicates as "being able to choose to do something only if it is consistent with demands one (or anyone) would make of anyone (hence that one would make of oneself) from a standpoint we can share as mutually accountable persons" (*Second-Person Standpoint*, 35). I will suggest that we don't *just* take a second person standpoint toward ourselves, though we can. I will also deny the Strawsonian assumption that "quality of will" is all that matters to us about ourselves, as reflected in our (fitting) self-reactive attitudes.
39. I have in mind contractualists such as T. M. Scanlon and Stephen Darwall (who explicitly locates himself in the Strawsonian tradition), who offer accounts of the construction of shared normative expectations about when it is right to hold people responsible. I am leaving aside disagreements amongst them (and other contractualists, not to mention other kinds of constructivists) about exactly how to conceive of the production of "what we owe to each other" (Scanlon, *What We Owe to Each Other*) or of "second-personal reasons" and the moral obligations to which they are intrinsically connected (Darwall, *Second-Person Standpoint*). For Darwall, "moral norms regulate a community of equal, mutually accountable, free and rational agents as such, and moral obligations are the demands such agents have standing to address to one another and with which

they are mutually accountable for complying" (*Second-Person Standpoint*, 101). For Scanlon, in the realm of "what we owe to each other" judgments of right and wrong are "judgments about what would be permitted by principles that could not reasonably be rejected, by people who were moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject" (*What We Owe to Each Other*, 4). Both of these accounts postulate idealized moral agents. My own position, which does not take the justification of principles by hypothetical, idealized beings to be sufficient for establishing their authority, is much closer to the more naturalized constructivist account developed by Margaret Urban Walker. The authority of normative expectations, for Walker, is found in the actual confidence that people have in them after subjecting them to what she calls "transparency testing"; norms that people remain confident in even when the norms are made transparent can be imbued with moral authority ("Seeing Power in Morality," 9; see also *Moral Understandings*, chapter 3).

40. Scanlon, *What We Owe to Each Other*, 4.
41. Notice that if fairness did not matter to us, then the fact that it was unfair to hold someone responsible for something would be the "wrong kind of reason" for concluding that they are not responsible.
42. Of course, we do have normative expectations according to which we can permissibly hold people responsible for what is beyond their control in the way that everything is beyond our control if determinism is true; that is—as Strawson points out—our ordinary interpersonal relationships depend on our being responsible without having libertarian freedom.
43. Williams, *Shame and Necessity*, 69.
44. Consider also Williams's famous remark that "One's history as an agent is a web in which anything that is a product of the will is surrounded and held up and partly formed by things that are not, in such a way that reflection can go only in one of two directions: either in the direction of saying that responsible agency is a fairly superficial concept, which has limited use in harmonizing what happens, or else that it is not a superficial concept, but that it cannot ultimately be purified—if one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for, one must accept much that makes its claim on that sense solely in virtue of its being actual" (*Moral Luck*, 29–30). Dr. Rojas has a self-reactive attitude in response to her own "impure" will. See also Walker, "Moral Luck and the Virtues of Impure Agency," for the idea of "impure agency."
45. Nagel insists on the possibility of an objectively real morality and would reject my metaethical position. Unlike Nagel, I take the objective standpoint—"the view from nowhere" (Nagel, *The View from Nowhere*)—to be a standpoint from which nothing matters, though from which we can see ourselves, from the outside, as beings to whom things matter, and from which we can discover non-normative facts that have implications for normativity.
46. Nagel, *Mortal Questions*, 166.
47. Nagel, *View from Nowhere*, 3.
48. Nagel, *View from Nowhere*, 3.
49. Lewis, *A Grief Observed*, 32.
50. Didion, *The Year of Magical Thinking*, 188–89.
51. The claim that the loss is bearable is not exactly a construal of the loss as bearable but rather an inference from the fact that the loss is being borne.

52. Frankfurt, *Importance of What We Care About*, 8.
53. This suggests a novel way of understanding why “ought implies can” is false in the case of unthinkable actions that one ought to, but cannot, avoid committing: when one has the subjective deontic experience of something as being unthinkable, the fact that one cannot avoid committing the unthinkable action is a reason of the wrong kind for concluding that the action is not required. “Cannot” does not imply “not the case that ought” because “cannot” is an extrinsic reason that does not bear on the question of whether the action is required.
54. See Tetlock et al., “The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals.”
55. Nagel, *Mortal Questions*, 172.
56. Walker, “Moral Luck and the Virtues of Impure Agency,” 18.
57. See D’Arms and Jacobson (*Rational Sentimentalism*, 174) for their view that onlookers admire the lorry driver’s attitude of guilt in a sort of aretaic judgment; however, D’Arms and Jacobson maintain that the guilt is admirable while being *unfitting*.
58. Note that what I am talking about would be a “transformative experience” according to L. A. Paul’s usage of the term; that is, the transformation I am referring to would be a personal transformation as well as an epistemic transformation. In an epistemic transformation one gains knowledge that one previously lacked, namely, knowledge of the subjective value of having some new experience; in a personal transformation it is “what it is like for you to be you” (Paul, *Transformative Experience*, 16) that changes, leaving one with a changed point of view, or changed preferences such that even an experience of the “same” type that one has had before would now be different. Paul uses the term “transformative experience” to refer to experiences that are at the same time both epistemically and personally transformative: “Having a transformative experience teaches you something new, something that you could not have known before having the experience, while also changing you as a person” (*Transformative Experience*, 17). Paul’s interest in transformative experiences is in the fact that they “raise a special problem for decision-making” (*Transformative Experience*, 18), namely, “not only do you not know the values before you’ve had the relevant experience, but having the experience can change your preferences, and so the values you would (*per impossibile*) assign these outcomes before having the transformative experience could be radically different from the values you’d assign to the relevant outcomes *after* having had the experience” (*Transformative Experience*, 32; italics in the original). My interest in transformative experience is somewhat different. Idealization that involves trying to imagine—and identify with the subjective standpoint of—a self who has undergone a transformative experience is complicated not only by the fact that we cannot know enough about this self, but also by the fact that because one would become an incommensurably different self through the transformation, the transformation will involve unique loss; this is what leaves us not only unsure (in our decision-making) but also ambivalent (in our emotional attitude) about undergoing the transformation or even claiming the subjective standpoint of the transformed self as our own.
59. As D’Arms and Jacobson analyze recalcitrance, “recalcitrance involves a failure of reasons responsiveness by one’s own lights. When one has both an emotion and a judgment that deems it unfitting, at least one of these states is not responding properly to one’s reasons” (*Rational Sentimentalism*, 67); they further clarify that the reasons responsiveness, both in the case of the emotion and in the case of the judgment, is a responsiveness to reasons of fit: “they are not both responding properly to reasons of the right kind: evidential reasons for belief, and reasons of fit for emotion” (*Rational Sentimentalism*, 67 n9). The reasons of fit I have been discussing are reasons of fit for *emotional* attitudes, namely,

attitudes that involve emotions that construe their objects as having certain normative features; in contrast, the attitude of belief does not construe its object as having any normative features, but rather as having the non-normative feature of being the case (so that if I have a belief that *p* then I take it that it is the case that *p*, or I take it that *p* is true). In my example of the self-aware phobic, I could say that this phobic has reasons of fit for a belief (that the snake cannot kill me) and reasons of fit for the emotional attitude of fear, and that these cannot both be good reasons of fit, so the attitudes cannot both be fitting. Then in addition to there being good extrinsic reasons to rid myself of the fear, there are also good reasons of fit for the (non-normative) belief that pulls in the opposite direction as the reasons of fit for the (normative) attitude of fear.

60. See D'Arms and Jacobson's similar point about a grieving widow ("Moralistic Fallacy," 77).
61. Proust, *Within a Budding Grove*, 340.
62. Marušić, *On the Temporality of Emotions: An Essay on Grief, Anger, and Love*, 93.
63. Frankfurt, *The Importance of What We Care About*.

## BIBLIOGRAPHY

- Brown, Cullin. "Two Levels of Response-Dependence about Responsibility." Unpublished manuscript. (n.d.)
- . "Subjectivism for Deeply Social Beings." Unpublished manuscript. (n.d.)
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (2000): 65–90.
- . "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgementalism)." *Royal Institute of Philosophy Supplements* 52 (2003): 127–45.
- . *Rational Sentimentalism*. New York: Oxford University Press, 2023.
- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.
- De Wijze, Stephen. "Tragic-Remorse—The Anguish of Dirty Hands." *Ethical Theory and Moral Practice* 7 (2004): 453–71.
- Didion, Joan. *The Year of Magical Thinking*. New York: Vintage Books, 2005.
- Frankfurt, Harry. *The Importance of What We Care About*. New York: Cambridge University Press, 1988.
- Hieronymi, Pamela. "The Wrong Kind of Reason." *The Journal of Philosophy* CII, no. 9 (2005): 437–57.
- . "I'll Bet You Think This Blame is About You." *Oxford Studies in Agency and Responsibility* Vol. 5: Themes from the Philosophy of Gary Watson. (2019): 60–87.
- . "Introduction" to *Minds that Matter*. Available at <https://hieronymi.humspace.ucla.edu/in-progress/>. (n.d.) Last accessed April 14, 2024.
- Hume, David. *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge. Oxford: Oxford University Press, 1978.
- Lenman, James. "Naturalism without Tears." *Ratio* 22 (2009): 1–18.
- Lewis, C. S. *A Grief Observed*. New York: Bantam Books, 1963.
- Marušić, Berislav. *On the Temporality of Emotions: An Essay on Grief, Anger, and Love*. Oxford: Oxford University Press, 2022.

---

Medecins sans Frontieres / Doctors without Borders. "Silent Wounds: Exploring the Moral and Ethical Challenges of COVID-19." March 11, 2022. <https://www.doctorswithoutborders.org/latest/silent-wounds-exploring-moral-and-ethical-challenges-covid-19>. Last accessed April 14, 2024.

Nagel, Thomas. *Mortal Questions*. New York: Cambridge University Press, 1979.

———. *The View from Nowhere*. New York: Oxford University Press, 1986.

Paul, L. A. *Transformative Experience*. Oxford: Oxford University Press, 2014.

Proust, Marcel. *Within a Budding Grove*. Translated by C. K. Moncrieff and Terence Kilmartin, and revised by D. J. Enright. New York: Modern Library, 1998.

Robinowicz, Wlodek, and Toni Rønnow-Rasmussen. "The Strike of the Demon: On Fitting Pro-attitudes and Value." *Ethics* 114, no. 3 (2004): 391–423.

Scanlon, T. M. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

Shoemaker, David. *Responsibility from the Margins*. New York: Oxford University Press, 2015.

———. "Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame." *The Philosophical Review* 126, no. 4 (2017): 481–527.

———. "Response-Dependent Theories of Responsibility." In *The Oxford Handbook of Moral Responsibility*, edited by Dana Kay Nelkin and Derk Pereboom, 304–24. Oxford: Oxford University Press, 2022.

Strawson, P. F. "Freedom and Resentment." In *Proceedings of the British Academy*. London: Oxford University Press, 1962.

Street, Sharon. "Constructivism about Reasons." *Oxford Studies in Metaethics* 3 (2008): 207–45.

———. "Coming to Terms with Contingency: Humean Constructivism about Practical Reason." In *Constructivism in Practical Philosophy*, edited by James Lenman and Yonatan Shemmer, 40–59. Oxford: Oxford University Press, 2012.

Tessman, Lisa. *Moral Failure: On the Impossible Demands of Morality*. New York: Oxford University Press, 2015.

———. *When Doing the Right Thing Is Impossible*. New York: Oxford University Press, 2017.

Tetlock, Philip, Orié Kristel, S. Beth Elson, Melanie Green, and Jennifer Lerner. "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals." *Journal of Personality and Social Psychology* 78, no. 5 (2000): 853–70.

Walker, Margaret Urban. "Moral Luck and the Virtues of Impure Agency." *Metaphilosophy* 22, nos. 1/2 (1991): 14–27.

———. *Moral Understandings: A Feminist Study in Ethics*. New York: Routledge, 1998.

———. "Seeing Power in Morality: A Proposal for Feminist Naturalism in Ethics." In *Feminists Doing Ethics*, edited by Peggy DesAutels and Joanne Waugh, 3–14. Lanham, MD: Rowman and Littlefield, 2001.

Watson, Gary. *Agency and Answerability: Selected Essays*. Oxford: Clarendon Press, 2004.

Williams, Bernard. *Moral Luck*. Cambridge: Cambridge University Press, 1981.

---

———. *Shame and Necessity*. Berkeley: University of California Press, 1993.

Wolf, Susan. *The Variety of Values*. New York: Oxford University Press, 2015.

Wong, David. *Natural Moralities: A Defense of Pluralistic Relativism*. New York: Oxford University Press, 2006.