

## I'll Bet You Think This Blame Is About You

May 15, 2018  
Pamela Hieronymi  
[hieronymi@ucla.edu](mailto:hieronymi@ucla.edu)

There seems to be widespread agreement, in the current philosophical literature, that what it is to be morally responsible for something is to be deserving of certain consequences on account of that thing.<sup>1</sup> It is a natural enough thought. We can call it the “merited-consequences” conception of responsibility. “Blame” is then used to refer to a large range of consequences that one might merit in cases of moral failure: judgment, criticism, sanctions, opprobrium, attitudes such as resentment or indignation, alterations in one’s interpersonal relationships, etc.

I have come to think there is something off, or askew, in thinking of moral responsibility in this way, though I find it hard to articulate clearly just what it is. It seems to me that the kind of moral responsibility the merited-consequences conception is trying to capture could be better captured by noting the characteristic way in which certain minds can rightly matter to other such minds—the way in which certain minds can carry a certain kind of importance, an importance that is made manifest in certain sorts of responses. Mattering, not meriting, is what seems to me central.

However, since I cannot yet better articulate the alternative, I will continue in the merited-consequences framework. I would like to focus on a particular class of the consequences one might merit, in response to moral failing: certain responses to moral failing are non-voluntary, in a sense I will explain below. The non-voluntariness of these reactions has two important upshots. First, questions about their justification will be complex. Second, because they are non-voluntary, they are

---

<sup>1</sup> Or, more exactly, to be responsible is to be such that certain consequences might be deserved or merited or appropriate. Recently the “meriting” or “appropriateness” relation has received some scrutiny. See, for example, the positions outlined in the introduction to Randolph Clarke, Michael McKenna, and Angela M. Smith, eds., *The Nature of Moral Responsibility* (New York: Oxford University Press, 2015), pp. 9–12. In addition to the notion of “basic desert” favored by Derk Pereboom, an idea of “fittingness” has been advanced by Michael Zimmerman, and an “alethic” relation by Gideon Rosen. See Derk Pereboom, *Free Will, Agency, and Meaning in Life* (New York: Oxford University Press, 2014); Gideon Rosen, “The Alethic Conception of Moral Responsibility,” in *The Nature of Moral Responsibility*, ed. Randolph Clarke, Michael McKenna, and Angela M. Smith (New York: Oxford University Press, 2015); Michael J. Zimmerman, “Varieties of Moral Responsibility,” in *The Nature of Moral Responsibility*, ed. Randolph Clarke, Michael McKenna, and Angela M. Smith (New York: Oxford University Press, 2015).

not well thought of as consequences voluntarily imposed upon the wrongdoer by the responder. By focusing on merited consequences and over-looking non-voluntariness, we risk misunderstanding the significance of moral criticism and of certain reactions to moral failure.

## 1. JUSTIFYING RESPONSES TO MORAL FAILURE

### 1.1 WHICH RESPONSES?

To start, let us consider responses to moral failure, and, for that, let us start with Gary Watson's landmark paper, "Two Faces of Responsibility."<sup>2</sup> That paper is a response to Susan Wolf's criticisms of what she calls "Real Self" views—views that say, roughly, that you are responsible for the outputs of your real self, simply because they reflect your real self.<sup>3</sup> Wolf made two criticisms: first, Real Self views do not account for the depth and force of moral criticism and, second, they do not take into account the unfairness of blaming someone who lacks the ability to become a better "real self." Watson replies by dividing responsibility into two faces: What he calls the "aretaic face" of responsibility concerns an evaluation of the real self: an evaluation of the agent and those choices attributable to them. Watson argues, against Wolf's first criticism, that negative aretaic evaluations do carry a kind of depth and force. He agrees that aretaic evaluations are not sensitive to concerns of fairness, but he suggests that this is as it should be, because concerns of fairness appear instead when considering the "accountability face," the aspect of responsibility that involves making demands, imposing sanctions, and what he calls "holding accountable."

I would like to consider the idea of a sanction in more detail—and, in the end, to make it more exacting than what Watson likely had in mind. To sanction someone is, as Watson agrees, to impose something unpleasant or unwanted upon that person in response to his or her violation of some norm, demand, or expectation. In holding accountable, he says, we "lay it down" that a

---

<sup>2</sup> Gary Watson, "Two Faces of Responsibility," *Philosophical Topics* 24, no. 2 (1996).

<sup>3</sup> Susan Wolf, *Freedom within Reason* (New York: Oxford University Press, 1990), chapter 2.

certain consequence will follow if the standard isn't met. (236–7) This seems true of sanctions. Sanctions are consequences that are *created* and *attached* to certain failures in certain contexts. Parents sanction or penalize children for misbehaving; society penalizes its members for violating civil law; nations impose sanctions on other nations for pursuing weapons programs; teachers penalize students for turning in work late. We even create and impose sanctions and penalties in games and sports.<sup>4</sup>

In contrast, a hangover is not a sanction for drinking too much. It is just a negative consequence. Similarly, the big mess in your kitchen is a negative consequence of last night's dinner party. The fact that you have to clean up that mess is not a sanction for having the party. It is just your job, and no one else's.<sup>5</sup> Sanctions are not simply negative consequences that no one else is required to bear for you. They are negative consequences that have been created or instituted and attached, by some person or body, to certain violations or shortcomings. As such, they trigger concerns of fairness. As Watson points out, "It is unfair to impose sanctions upon people unless they have a reasonable opportunity to avoid incurring them." (237)

How, then, are sanctions justified, morally? In *What We Owe to Each Other*, T. M. Scanlon provides a compelling answer: They are justified in the way actions are usually justified. Or, rather, according to Scanlon, actions are usually justified by appeal to contractualist reasoning, and this is also the way that both the institution and the imposition of a sanction will be justified.<sup>6</sup> But one need not accept the details of Scanlon's contractualism to see the appeal of understanding the

---

<sup>4</sup> Note that, so understood, sanctions need not have any distinctively moral significance. I will consider the idea that blame is a sanction imposed within the social institution called "morality" later.

<sup>5</sup> It is what Scanlon would call a "substantive responsibility"—a question of where obligations and burdens fall. I am, here, distinguishing sanctions from other (or mere) substantive responsibilities.

<sup>6</sup> See T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), chapter 6. In the context of that work, sanctions are instances of "substantive responsibilities:" benefits, burdens, and risks to be distributed according to ordinary contractualist reasoning. For Scanlon "substantive responsibilities" sharply contrast with what Scanlon there called "responsibility as attributability" and now calls "moral reaction responsibility." (For the latter, see "Forms and Conditions of Responsibility," in *The Nature of Moral Responsibility*, ed. Randolph Clarke, Michael McKenna, and Angela M. Smith (New York: Oxford University Press, 2015).

justification of *creating* a sanction to be, broadly speaking, a question about the creation or modification of a kind of social institution and the justification of *imposing* a sanction to be the justification of an ordinary action, in that institutional context. As noted by Watson, sanctions will be subject to concerns about whether their target had an adequate opportunity to avoid them, and so subject to concerns of fairness, but, we can now say that the notion of “adequate opportunity” will be constrained by the realities of the situation and the competing needs and interests of everyone affected. A student who claims they lacked fair opportunity to avoid the late penalty on their paper because the world is governed by deterministic causal laws will be disappointed. The justification of a sanction is, so to speak, political or institutional, not metaphysical.

Notice that, if we could maintain the Watsonian division into two faces, then the justification of responses to moral failing would be relatively straightforward—or, at least, it would be relatively clear how we should proceed in our theorizing. The first face concerns evaluations. An evaluation is justified in the way a belief is—roughly, by appeal to evidence and standards of epistemic reasoning. Philosophers think we know something about evidence and standards of epistemic reasoning. If we could maintain the division into two faces, then anything beyond mere evaluation would be understood as a kind of sanction, to be justified in the way that political or institutional structures and the actions within them are justified. And, again, we, as philosophers, have some thoughts about the justification of political and institutional structures and actions within them. And so, if we could maintain this division, we would have, available to us, two familiar frameworks for thinking about the justification of responses to moral failing.

## 1.2 BEYOND TWO FACES

Unfortunately, though, two faces are too few.<sup>7</sup> Mere evaluation does not fully capture blame, but the remaining responses are not all well characterized as sanctions—at least not once we understand

---

<sup>7</sup> See, e.g., Angela M. Smith, "Moral Blame and Moral Protest," in *Blame: Its Nature and Norms*, ed. Justin Coates and Neal Tognazzini (New York: Oxford University Press, 2013). For a three-fold division, see David Shoemaker, "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics* 121 (2011).

sanctions in the more exacting way I have suggested. And so we cannot avail ourselves of existing theorizing about epistemic justification, on the one hand, and the justification of creating and imposing sanctions, on the other. Instead, the justification of responses to moral failing must involve the messy middle ground between belief and action.

The remaining responses to moral failing include the “reactive attitudes” that P. F. Strawson put at center-stage: resentment, guilt, indignation, gratitude, trust, admiration.<sup>8</sup> Scanlon expands the category of “reactive attitude” to include such things as “withdrawal of trust” and a “decreased readiness to enter into special relations of friendship,” or to help a person with his projects, or to take pleasure in things going well for the person and to feel sad or regretful when they do not.<sup>9</sup> Others (prominently Michael McKenna and more recently Miranda Fricker) have noted the importance and role of communication in holding responsible.<sup>10</sup> We might also consider the possibility of a kind of moral judgment or condemnation, in a sense that goes beyond mere belief or evaluation, or what R. Jay Wallace calls “opprobrium”.<sup>11</sup>

The important point, for present purposes, is that many (though not all) of these remaining responses are non-voluntary, in a sense I am about to explain. Moreover, they *must* be non-voluntary, in this way, to play the roles they play and bear the significance they bear in our lives.

---

<sup>8</sup> See Peter F. Strawson, “Freedom and Resentment,” in *Free Will*, ed. Gary Watson (Oxford: Oxford University Press, 2003).

<sup>9</sup> Scanlon, “Forms and Conditions of Responsibility,” 92.

<sup>10</sup> Michael McKenna, *Conversation and Responsibility* (New York: Oxford University Press, 2012). Miranda Fricker, “What’s the Point of Blame? A Paradigm Based Explanation,” *Noûs* 50, no. 1 (2016).

<sup>11</sup> Watson is, I think, the first to break contemporary ground on the rich topic of judgmentalness. See Gary Watson, “Standing in Judgment,” in *Blame: Its Nature and Norms*, ed. Justin Coates and Neal Tognazzini (New York: Oxford University Press, 2013). For opprobrium, see R. Jay Wallace, “Dispassionate Opprobrium: On Blame and the Reactive Sentiments,” in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. R. Jay Wallace, Rahul Kumar, and Samuel Freeman (New York: Oxford University Press, 2011).

### 1.3 THE NON-VOLUNTARY<sup>12</sup>

The words “voluntary” and “involuntary” can be put to a variety of different uses, each important. To locate the sense currently at issue, let us start with what might seem to be a kind of spectrum between the clearly active and the clearly passive.<sup>13</sup> At the clearly active end stand ordinary actions, like raising your right hand, planting azaleas, or running for office. These seem clearly things you do. At the other extreme stand the (relatively) clear cases of passivity. These include not only being blown by the wind, succumbing to disease, and winning the lottery, but also having a headache, hearing a ringing in your ears, or seeing spots. These are things you did not choose and are not up to you. Between these extremes, it seems, stand an interesting class of states of mind: states which appear passive, when contrasted with ordinary intentional actions, but which seem active, when contrasted with sensations or perceptions.

The starkest example is believing. Believing is importantly unlike raising your right hand or running for office, in that you cannot believe “voluntarily” or “at will,” in the following specific sense: While you can raise your right hand or run for office for any reason that you think shows it worth doing—to win a bet, or make a joke, or make a point—you cannot believe something (e.g., that the butler did it) in order to win a bet, make a joke, or make a point—even if you think it would be worth doing. You can only believe what you take to be true. You are constrained, in your believing, in a way that you are not, in your acting. Thus, compared to action, believing seems passive.

On the other hand, believing is importantly unlike having a headache or seeing spots. What you believe, what you take to be true, is, in a certain way, up to you. If you think appearances are deceiving, you will not believe your eyes. If, upon considering the evidence, you change your mind

---

<sup>12</sup> Some material from this section is repeated, with minor modifications, in “Reasoning First,” in *Routledge Handbook of Practical Reasoning*, ed. Ruth Chang and Kurt Sylvan (in process). It will appear in *Minds that Matter* (in progress).

<sup>13</sup> It is useful, rhetorically, to start with the idea of a spectrum. In the end, it should be rejected. Thanks to Richard Moran for pressing this point.

about the butler's guilt, you will no longer believe him guilty. Our beliefs reflect or embody our thoughts—in particular, our thoughts about what is so. And thinking is something we do. As thinkers, we do not simply suffer or experience our beliefs, in the way we suffer a headache or experience a visual illusion. Our beliefs are up to us in the way that our answer to a question is up to us—where, by an answer to a question, I do not mean an oral or written response to a question (which is an ordinary action), but rather the resolution of a question arrived at in one's own mind, that which an oral or written response might or might not sincerely express. When you answer a question, your answer is not simply out of your control. You may arrive at it with some care. And your answer is up to you: it will change with your changing assessment of the subject matter. Although you are constrained, in your answer, by your own assessment of the subject matter, it is your own assessment that determines your answer.

Moreover, the assumption that our beliefs are up to us in this way makes sense of our practice of asking others to defend their beliefs, to provide us with *their* reasons for believing. If we ask someone why she has a headache or hears a ringing in her ears, we expect an account of what produced this state of discomfort. But if we ask someone why she believes the butler did it, we expect to be given, not an explanation of the history of a particular mental episode, but something quite different: a present case for the butler's guilt, a case to *support* the belief. That is to say, we accept, as an explanation for her state of mind, considerations that bear on a very different subject matter; we accept, as an explanation for the psychological state of the person standing before us, considerations that bear on a subject matter removed in space and time—facts about the butler, his access to the home, his motives, etc. Moreover, if we think the reasons the person provides are inadequate, we will criticize the person for their state of mind—we will think they ought to revise it. We might, in addition, offer to them similarly remote reasons and expect the person to respond by

forming a different, or opposing, belief. All of this would be unthinkable, if we did not understand her state of mind as up to her in the way her answers up to her.<sup>14</sup>

So, belief is not active in the way an ordinary intentional action is, it is not voluntary, but it is not passive in the way that perception and sensation are—nor is it involuntary in the way digestion or blinking or accidents are. It is up to you in the way your answer to a question is up to you. Thus I will say that believing is *non-voluntary*. It is, I think, a non-voluntary activity. Many other states of mind—indeed, many of those we find most important—are also non-voluntary. I would include, in this wide category, intentions, decisions, and attitudes such as resentment, gratitude, trust, admiration, contempt, and satisfaction at a job well done. These are, like belief, attitudes for which we can be asked our reasons, but which we cannot adopt at will.

I would characterize these non-voluntary attitudes as forms of question-answering—it is this form, I think, that gives applicability to the request for one's reasons. In saying this, in saying that to believe *P*, for example, is to answer the question of whether *P*, or that to intend to *x* is to answer the question of whether to *x*, I do not mean to posit a new, independent psychological event or activity, the answering of a question, that somehow accompanies believing or intending. Rather, I mean to claim that belief and intention, themselves, are helpfully thought of as question-answerings; question-answering is something like a genus into which these attitudes fall as species.

If we see these activities as question-answerings, we can say why they are not voluntary in the sense specified: your reasons for them will be the reasons you take to bear on the relevant question. You will believe *P* for reasons you take to show that *P*; you will intend to *x* for reasons that you take to bear sufficiently on whether to *x*. But, of course, you might take yourself to have *other* reasons: perhaps you could get a good night's sleep if only you could believe everything will be okay. You

---

<sup>14</sup> The point is made by Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001). Note, too, we will criticize the person if the reasons given are inadequate. We might present other, similarly remote, considerations and expect the person to form some other, or contrary, belief in response. This would, again, be unthinkable, if we did not understand the state of mind as in some way an answer to the question on which those reasons bear.



cannot believe everything will be okay for this reason, because you cannot answer a question for reasons you do not take to bear on it: If you were to answer the question for that reason, you would have *therein* taken it to bear on the question.<sup>15</sup> Thus, belief is non-voluntary.<sup>16</sup>

It turns out the same is true of intention. You might have reason that you take to be sufficient reason to *intend* to *x*—reason enough to house the intention—but that you do not take to be reason enough to *x*—not reason enough to act. Perhaps you have no intention of marrying your partner, and they are very unhappy about this fact. Because you like to please your partner, you would be happy to house the intention—so long as you do not need to actually go through with the marriage. You are out of luck. In order to intend to marry, you have to decide to marry—to intend, you must answer the question of whether to *act*, not just the question of whether to intend. And so, even though you take yourself to have reason enough to intend, you will not be able to intend. Somewhat surprisingly, although you can *act* at will—though you can act for any reason you take to show the action sufficiently worth doing—you can no more intend at will than you can believe at will. While *actions* are voluntary, *intentions* are not.<sup>17</sup>

Notice that attitudes that are non-voluntary in this specific sense, are also, and therefore, subject to a wrong-kind-of-reasons problem: you might find yourself with reasons that you take to show them worth having that you do not take to bear on the relevant questions. Getting a good night's

---

<sup>15</sup> Importantly, to take a consideration to bear on a question is not to form a belief about the consideration, the question, and the “bearing on” relation. It is, rather, to employ the consideration in addressing the question.

<sup>16</sup> I have argued that you cannot believe at will in Pamela Hieronymi, "Controlling Attitudes," *Pacific Philosophical Quarterly* 87, no. 1 (2006); "Believing at Will," *Canadian Journal of Philosophy, Supplementary Volume* 35 (2009).

<sup>17</sup> The argument that you cannot intend at will appears in "Controlling Attitudes." The marriage example appears in "Responsibility for Believing," *Synthese* 161, no. 3 (2008); "Reflection and Responsibility," *Philosophy and Public Affairs* 42, no. 1 (2014); "Forgiveness, Blame, Reasons..." in *3am: magazine*, ed. Richard Marshall (2013).

rest is the wrong kind of reason for believing everything will be okay. The same sort of problem can be generated for any member of this class.<sup>18</sup>

The non-voluntariness of these attitudes can seem to be an affront to our powers—as though there is something we cannot do, because of some shortfall in our abilities. But there is no shortfall, any more than our inability to draw a square circle represents a shortfall. The illusion of shortfall is generated by the fact that we are both rational and reflective: we can both answer questions and think about our own answers. We exercise our rational agency by finding reasons sufficient—that is, by answering questions and therein forming attitudes. There is no shortfall here: We can consider any question of which we can conceive, and we can answer it for any consideration that we take to bear sufficiently on it. By answering certain questions, we therein form certain attitudes. Because we are also reflective creatures, we can also think *about* our attitudes and notice that they are costly, or inconvenient, or that other attitudes would better serve our purposes. In many cases, we thereby encounter reasons of the wrong kind: reasons that count in favor of changing our attitudes that we do not take to bear, or bear sufficiently, on the relevant question. We then find ourselves with reasons for something that is itself the kind of thing done for reasons, and yet unable to do it for the reasons we have. But this inability represents no shortfall: the thing we are unable to do is not, itself, a coherent possibility: you cannot answer a question for reasons that you do not take to bear

---

<sup>18</sup> I am understanding the wrong-kind-of-reasons problem in the way I have elsewhere, and I will use “wrong kind” accordingly. For a short summary, see “The Use of Reasons in Thought (and the Use of Earmarks in Arguments),” *Ethics* 124, no. 1 (2013).

It is difficult to generate the problem for intention, because there are very few constraints on the reasons for which one can act (most any consideration could, in principle, bear on the question of whether to  $x$ ) and it is possible to act as a way of making yourself intend. (If you are unhappy that I have no intention to attend your party, I can decide to attend your party in order to keep you happy—even if what you really care about is my intention, not my attendance.)

It may be worth noting that, in the marriage example, pleasing your partner is not exactly the wrong kind of reason for intending, because, if you thought that housing the intention were reason enough to marry, you could decide to marry in order to have the intention. The reason *bears* on the question, but you do not take it to be sufficient reason to settle the question. In contrast, the Toxin Puzzle case (Gregory Kavka, “The Toxin Puzzle,” *Analysis* 43 (1983).) and the original case of Mutual Assured Destruction are ones in which the reason to intend does not even bear on the question of whether to act, because the reason to intend disappears before the time of action, and this is known in advance. These cases present reasons that are genuinely of the “wrong kind”.

sufficiently on that question. Why not? Because, if you answer a question for a reason, you will *therein* have taken the reason to bear sufficiently on that question. Being disappointed that we cannot answer a question for a reason we do not take to bear sufficiently on it is like being disappointed we cannot draw a square circle. Our feeling of lack is an illusion generated by the fact that we are able both to answer questions and to reflect upon our own answers.<sup>19</sup>

Thus the fact that an attitude is non-voluntary, in the sense I have outlined, should not lead us to think it is not up to us or that we are passive with respect to it. These attitudes are up to us in the way our answers to questions are up to us—they express our take on things. Unlike your bone structure or birthplace, they change with your changing evaluation of what is true, good, worthwhile, or important. Unlike your headache or a visual illusion, it makes sense to ask you for your reasons for them. They are not simply out of your control, even though you cannot adopt them for any reason you take to show it worth doing.

---

<sup>19</sup> This point appears in Hieronymi, "Believing at Will," p. 173. The editors raise a helpful objection: Grant that we cannot settle a question for reasons we do not take to bear on it. Our disappointment need not be understood as the disappointment of being unable to do this incoherent thing. Our problem (pleasing our partner, getting some sleep) would be solved if only we could assess the reasons differently—if we could take them to bear on the question sufficiently. But there is nothing incoherent in the possibility of assessing the reasons differently. Someone else may take their partner's desire for an intention to marry to be reason enough to marry, or someone else may think they are a kind of anti-clairvoyant, such that their own anxiety is evidence that everything will work out. These are coherent possibilities, and, if we were like these people, we would not face our problem. Thus, it seems, our disappointment lies in a genuine shortcoming: we cannot bring ourselves to be like these people, we cannot bring ourselves to assess the reasons in this way, and so cannot enter the desired state of mind.

In reply, I will first note that whether you can *bring yourself* to assess the reasons in a different way has not been considered. Sometimes you will be able to do so, sometimes you will not. Perhaps, by following Pascal's advice, you can come to see your partner's concern as reason enough to marry, or perhaps, by establishing side bet with your friend, you can come to see the prize money as reason enough to drink the toxin. Being unable to bring yourself to assess the reasons differently is a genuine shortcoming, one you might, or might not, face, and facing it might, or might not, disappoint you. (But, even if you are able to bring yourself to assess the reasons differently, you would not, thereby, have believed, or intended, at will or voluntarily. Cf. "Controlling Attitudes.")

The objector might restate the objection. Our disappointment lies in a genuine shortcoming: We are unable (now) to assess the reasons differently than we do. But if this means "...differently than we do assess them, however that may be" then we have again reached an incoherence. If it means, instead, "...differently we do assess them, namely, we are unable to take our partner's desire to be reason enough to marry," then it is, I think, false—you *could* assess the reasons in that way. But you do not: you are committed to a different assessment. (More needs to be said, on this point.)

#### 1.4 NON-VOLUNTARY REACTIONS TO MORAL FAILURE

Returning, now, to the main line of thought, armed with this idea of the non-voluntary: we considered Watson's division into two faces, we made the idea of a sanction a bit more exacting, and then we noticed that not all responses to moral failing can be categorized as either evaluation or sanction. I claimed that some of the remaining responses are non-voluntary, in the sense just explained, and, moreover, that they must be non-voluntary, in this sense, to play the roles they play and bear the significance they bear in our lives.

To illustrate the last claim, consider trust and pride. Trusting (in the sense of having great confidence in) my children, or taking pride in their accomplishments, builds their self-esteem and aids their personal development. Even so, I do not—and could not—feel great confidence in them or take pride in their accomplishments in order to build their lagging self-esteem. If I attempted to do so, my “pride” or “trust” would be inauthentic and so, most likely, ineffective. To be sure, I can *express* my genuine trust or *voice* my genuine pride in order to build their self-esteem.

Communication is voluntary.<sup>20</sup> But trust and pride, themselves, are not—they have their own reasons, and aiding personal development is not among them. Moreover, this must be so, if they are to play the roles they play and bear the significance they bear in our lives.

So, too, with the reactive attitudes Strawson put at center-stage—resentment, gratitude, indignation, trust, distrust, admiration, contempt. These do not, and could not, reflect a person's all-things-considered judgment about the best response to have in the situation—as they could, and typically would, if they were voluntary.<sup>21</sup> They reflect a person's take on only certain aspects of their situation. And that must be so, for them to play the roles they play and bear the significance they bear in our lives.

---

<sup>20</sup> Or, it is often voluntary.

<sup>21</sup> “Typically,” because of cases of weakness of will. Voluntary responses reflect a person's take on the question of whether to *x*, where that question *ought to be* answered by the same range of considerations as the question of whether *x* would be the best response all-things-considered (but sometimes it is not so answered). In contrast, non-voluntary reactions reflect a person's take on a narrower range of considerations.

But, because these attitudes are not sensitive to every aspect of a person's situation, a person cannot be justified in them by appeal to just any aspect of the situation, nor, importantly, can their justification be challenged by just any aspect. These attitudes embody your take, not on the all-things-considered question of whether it would be best to have those attitudes, but instead on some narrower set of questions which concern (not the attitude itself, but rather) only certain aspects of your situation. And so the question of whether you have reason to resent, e.g., and therefore the question of whether you are justified in resenting, cannot be answered by appeal to just any aspect of your situation or relationship. It can be answered only by appeal to a narrower, more specific set of considerations.<sup>22</sup>

Exactly what the relevant aspects are, for a given attitude, is a difficult but worthwhile question. To illustrate, consider resentment. We can distinguish between reasons for which one can, so to speak, directly revise one's resentment and other reasons that count against resenting.<sup>23</sup> A legitimate excuse, something that shows that in fact no disregard was shown, is a reason for which you can directly revise your resentment. In fact, if you do not, you are being irrational. In contrast, the fact that your resentment is ruining your own happiness is not a reason for which you can directly revise your resentment—it is, instead, a reason to take action in order to attempt to rid yourself of your resentment: to undergo therapy or take medication or at least put some effort into trying to reframe or reinterpret your situation. Notice that the attempt to reframe or reinterpret the situation is, itself, an effort to find the right kind of reason—reasons for which you can directly or rationally revise your resentment. The negative effect on your happiness is a reason to engage in self-management. It is not a reason that is internal, so to speak, to resentment—it is the wrong kind of reason for the attitude itself.

---

<sup>22</sup> Though I am using the notion of “question-answering,” at this point, I do not believe it is essential to the point.

<sup>23</sup> I would cash out the metaphor of “directness” (and of what is “internal”) in terms of questions. (Cf. footnote 38.)

Because a person's reasons for resenting do not concern her own happiness, facts about her happiness will neither show her justified in resenting nor call the justification of the resentment, itself, into question.<sup>24</sup> Facts about her happiness will justify, or call into question, other activities.<sup>25</sup>

To understand what would either justify resentment or call it into question, we need to identify the right kind of reason for it. Elsewhere I use a puzzle about forgiveness to argue that the reasons proper to resentment might bear on one of four questions: whether a wrong was done; whether you have a claim not to be thus wronged; whether the wrongdoer is a moral peer; and whether the wrong done poses an on-going threat to your moral standing.<sup>26</sup> But whether or not one agrees with this specific account, if one accepts that only certain reasons are proper, or internal, or of the right kind, one will need to heed this difference, when thinking about the justification of the attitude. Doing so will complicate our assessment of their justification.

### 1.5 CASE STUDIES

To illustrate, I will consider three case studies, each drawn from Scanlon's recent paper "Forms and Conditions of Responsibility."

To begin, consider Scanlon's treatment of moral evaluation, Watson's aretaic face. Considering its justification, he says,

If the reactive attitude is simply moral evaluation... there is a question of what reason we have for being concerned with other people's moral quality. This concern might seem like pointless grading unless it serves some practical or prudential purpose. (92)

---

<sup>24</sup> The same can be said about concerns about the wrongdoer's opportunities to avoid being the target of resentment. See "The Force and Fairness of Blame," *Philosophical Perspectives* 18, no. 1 (2004).

<sup>25</sup> I am relying on a principle: Strictly speaking, the reasons by which an individual can be justified or criticized for a rational activity (such as believing, intending, resenting, trusting, being proud) must be the kind of reasons for which one might engage in that activity (they must be the right kind of reason, bearing on, or being taken to bear on, the right question). If, instead, you criticize or justify someone by appeal to other reasons (your resentment is making you so miserable; you should give it up), either your criticism is simply inapt or else you are really advocating, criticizing, or justifying some nearby activity of self-management. This seems foundational. It seems to me one grain of truth in "ought-implies-can." Thanks to the editors for asking for clarification.

<sup>26</sup> "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, no. 3 (2001). See also footnote 34, for the idea of protest.

Scanlon accepts “pointless grading” as a possible criticism of moral evaluation. He thinks you might rebut the criticism if you have good reason to evaluate others, and he hopes to show that concern for the nature of your relations to others provides such reason.

My concern with Scanlon’s position is subtle but, I think, crucial. It is this: the charge of pointlessness cannot be leveled at the evaluation *itself*, but must apply to some other activity. If, in the course of a meeting, you repeatedly interrupt our colleague in an overtly rude way, I have adequate reason to conclude that you have been rude. I cannot be charged with pointless grading simply because I form that belief. Moreover, the problem with charging me with pointless grading is not that it would be unfair or objectionable to criticize me for doing something I cannot avoid doing. The problem is deeper: it is that beliefs do not have points. Beliefs are not formed or held for a purpose, and so the charge of pointlessness cannot attach to the believing itself, but must rather attach to something else—something like patterns of attention or activities of investigation.<sup>27</sup> If I sit down to re-grade last year’s homework assignments, it is the sitting down to grade, the spending my afternoon in that way, that is objectionably pointless. My belief that a particular assignment merits a C+ also has no point, but that is no criticism of it.

We can be more charitable to Scanlon and run into the same problem. By “pointless grading” Scanlon is not meaning to level the charge I would face if I were to sit down with last year’s homework. I suspect he has in mind, rather, the charge of a certain kind of busybodiness or of indulging in a kind of connoisseurship with respect to our fellow humans and their choices. But even so, connoisseurs of moral quality are objectionable, not simply because they form accurate opinions about the quality of others’ wills, but because they go out of their way to consider their quality, make the most of any occasion for evaluation, and, perhaps most objectionably, take some

---

<sup>27</sup> The charge of judgmentalness seems to attach to some attitude which might accompany the belief. But exactly what attitude is an interesting and difficult question. See Watson, “Standing in Judgment.”

sort of pleasure or comfort in so doing. But these activities and attitudes go beyond the accurate opinion. Our criticisms must be matched to the non-voluntary attitude in question.

The same problem appears in another case Scanlon considers, from a George Elliot novel. In the novel, Miss Asher shares a meal with her suitor, Captain Wybrow. She is off-put when he asks if she would like some jelly. Asher thinks that, by now, Wybrow should have noticed that she never takes jelly. This case entered philosophical discussion when Angela M. Smith used it to support of her view that we can be responsible for failures of attention.<sup>28</sup> John Martin Fischer and Neal Tognazzini objected, claiming that Wybrow cannot be responsible for his failure of attention unless it can be traced to something he freely did or freely omitted.<sup>29</sup> Scanlon defends Smith, noting first that, if we consider positive responses, such as continued interest, the tracing requirement is implausible: Asher need not consider whether Wybrow freely made himself attentive before responding to his attentiveness positively. Scanlon then considers whether the tracing requirement should apply to negative reactions, reactions that impose a loss on their target.

Scanlon accepts the possibility that a negative response from Asher could be shown unjustified by a loss to Wybrow, but he thinks that, in this case, as a “first-order, moral” matter, it is not.<sup>30</sup> He says, “It is more than can be asked of Miss Asher that she should accept an unconcerned and inattentive suitor as long as he has not freely chosen to be such a person.” (102)

---

<sup>28</sup> Angela M. Smith, "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115, no. January (2005): pp. 242ff.

<sup>29</sup> John Martin Fischer and Neal Tognazzini, "The Truth About Tracing," *Noûs* 43, no. 3 (2009).

<sup>30</sup> Scanlon explains his methodology in the paragraph that spans pp. 98–99. He says “the particular reactive attitudes... can be made appropriate simply by what a person’s normative attitudes in general are like.” This sounds like the claim I am hoping to make: the right kind of reasons for the reactive attitudes are facts about the quality of the other person’s will. However, as Scanlon clarifies his claim, it seems he thinks that whether an attitude is “made appropriate” is “a first-order moral thesis about when we do and when we do not owe it to others not to modify our attitudes towards them in these ways.” If, by “first-order moral thesis,” Scanlon means “a thesis to be established by ordinary contractualist reasoning,” then I disagree: insofar as these reactive attitudes are non-voluntary, their justification will be insensitive to many of the considerations to which contractualist reasoning is usually open.



This seems to me the wrong thing to say. If we were considering the justification of a voluntary action, such as *marrying* Wybrow, or picking him up from the airport, then what Scanlon says would make sense: we would be confronting a “first-order, moral” question, and a loss to Wybrow would be, in principle, relevant. We could ask whether, in light of some loss to Wybrow, Asher could be required to marry him, and we could give an immediate answer: it is too much to ask.<sup>31</sup>

But if we are considering, instead, withdrawing interest or being off-put, then the loss to Wybrow seems irrelevant. It seems to be the wrong kind of reason. Asking Asher not to be off-put by Wybrow’s inattention, because of some loss he incurs, is like asking her not to resent Wybrow for disrespecting her, because it damages his self-esteem; or asking her to find him the most handsome man around, because it would improve his mood; or asking her to believe something false, because it would help everyone get along. To say it is *too much* to ask undersells the problem. It is an absurd thing to ask. One might say, modifying a complaint from Bernard Williams, that it is an assault on Asher’s integrity—it asks her to do something she has no reason to do.<sup>32</sup>

But, again, if the loss to Wybrow is the wrong kind of reason for Asher to revise her response, then the justification of the response, itself, is not called into question by it.

Of course, once we move from non-voluntary attitudes, such as belief or resentment or being off-put, to their expressions or the communication of such attitudes, then we are, at least often,

---

<sup>31</sup> In conversation Scanlon clarified that he meant, by “accept an unconcerned and inattentive suitor,” marriage. Thus my criticism does not apply to his intended meaning.

<sup>32</sup> Williams was concerned about cases in which a person has no reason to do what they are required to do because of contingent features of their “motivational system.” I am focusing, instead, on cases in which a person is being asked to respond to a reason of the wrong kind—a reason that does not bear on the right question to enable the response imagined. In fact I think that Williams’ criticisms of both utilitarian and Kantian moral theories can be well understood if one interprets Williams as seeing the theories as relying on the wrong kind of reason for virtuous actions. (Virtuous actions, unlike ordinary voluntary actions, are subject to something like a wrong kind of reason problem.) See Bernard Williams, “A Critique of Utilitarianism,” in *Utilitarianism: For and Against*, ed. J. J. C. Smart and Bernard Williams (Cambridge: Cambridge University Press, 1973). and “Internal and External Reasons,” in *Moral Luck* (1981). For an argument that Williams’ concerns are grounded in his internalism, see Pamela Hieronymi, “Internal Reasons and the Integrity of Blame,” (1996). website. For the suggestion that reasons of the wrong kind underlie Williams’ criticisms see “Virtue and Its Imitation” (Harvard University, 2000), chapter 5.

considering voluntary actions. Speaking, e.g., is a voluntary action; one can choose not to speak or choose to speak insincerely. As with any voluntary action, the justification of speech can be questioned in light of burdens on others. In fact, Scanlon's strategy for justifying moral reactions, by appeal to the importance of the quality of a person's will and the norms governing the specific relationship, seems to me correct, when applied to voluntary expressions and communications. I would only add this observation: once we recognize that the attitudes expressed are, themselves, non-voluntary, we have available an additional set of reasons by which to justify their voluntary expression: considerations of integrity, honesty, or authenticity.

I would like to consider one last set of examples from Scanlon. He considers both our responses to children and our responses to psychopaths. He notes that we do not properly resent the wrongs done to us by children, even in cases in which the child fully understands both the consequences and the significance of their action—that is, even in cases in which the child displays genuine ill will or malice. He says this is because our relationship with children is not one of moral equality but rather one of nurturing and care-giving. As he puts it,

our reactions to children are limited by the duties of care that are part of our relations with them... our proper responses to children's misbehavior are dominated by the aim of promoting their moral improvement. So the appropriate response to thoughtless or insulting behavior by a child... should mainly be guided by the aim of helping the child to learn how to do better.... (104).

While I agree that resenting a small child for even extreme ill will is inapt, I am not convinced it is our duties or aims with respect to the child's moral improvement that explains why. My doubts are brought out by Scanlon's treatment of the psychopath.

Scanlon thinks the ill will of a psychopath rightly elicits resentment, but the fact of mental illness generates duties that require us to act in ways other than our resentment might incline us. He notes that, unlike in the case of children, our resentment of psychopaths is not undermined by these duties, but simply in tension with them. But now we encounter a puzzle: why do our duties undermine resentment in one case and not in the other?

I suspect it is because it is not our duties or aims that explain the inappropriateness of resenting small children. When resentment is undermined by the fact of childhood, it must be because one has revised one's take on one of the questions or commitments internal to resentment, and I doubt that aims of nurture or duties of care are relevant—as is shown, I think, by the case of the psychopath. My own, controversial, opinion, built on my own analysis of resentment, is that we do not resent the ill will of small children because we do not see their ill will as a threat to our standing. In contrast, ill will from an adult can be a threat, even if that adult is suffering from a disease such as psychopathy or narcissistic personality disorder. If the actions and choices of such adults have not been in some way marked or contained by society—if the psychopath or narcissist is running a corporation or a country, for example—they are threatening and remain a proper object of resentment. But, again, this opinion is controversial. The important point is that our duties, or facts about the nature of our relationships, will not show a reaction such as resentment inapt unless they undermine the reasons proper to it.

I hope these case studies illustrate the way in which the justification of a non-voluntary reaction, such as resentment or indignation or distrust, will be more complicated than the justification of a voluntary action, such as speaking or marrying or sanctioning. The justification of a non-voluntary attitude must be constrained and informed by the nature of the attitude itself, taking into account the difference between the right and the wrong kind of reasons for it.

### **1.6 FURTHER QUESTIONS OF JUSTIFICATION**

But a large issue remains outstanding. We can grant that the reactive attitudes are not voluntary, and we can grant that they are not sanctions, but, even so, they are still our activities—they are attitudes for which we can be asked our reasons, manifestations of our take on the world, unlike our bone structure or our birthplace. They are also, in *some* sense, our collective creation, a part of our culture. And so their nature is not simply beyond or outside questions of justification.

On the other hand, we are not able, either individually or collectively, to immediately or directly revise the nature of these reactions—any more than we would be able, either individually or collectively, to immediately or directly change the meaning of a word or alter the significance of a gesture or a flag. This is another way in which they differ from sanctions or penalties—and also from rules of a game or positive law.

Still, to say that we cannot immediately or directly revise them is not to say that they cannot be revised—we might, together, take actions that change them. And so, again, they are not entirely beyond criticism.

At this point let us consider, very briefly, a recent suggestion by Miranda Fricker.

Fricker presents a very different way of thinking about the justification of blame.<sup>33</sup> She says, roughly, that blame may be justified if it is justified in its paradigm case, which she calls “Communicative Blame.” Communicative Blame is a specific form of “second-personal interaction,” in which one person “finds fault” with another and communicates the judgment of fault to him or her “with the added force of some negative emotional charge.” (172) Fricker claims that this form of exchange has an aim: to bring the wrongdoer to “feel sorry for what they have done,” in particular, to feel “remorse.” In the course of the typical exchange, either the wrongdoer is led towards remorse or the blamer is led away from finding fault. Either way, the exchange functions to achieve what Fricker calls “increased alignment of moral understanding.” It may also serve to shift the motivations of the parties into greater alignment. Insofar as its paradigm case works towards these worthy aims, Fricker thinks blame can be vindicated.

I would have serious misgivings about Fricker’s view if it required the blamer to adopt these aims—if it required the blamer to aim at remorse or at increased alignment of moral understanding or motivation. However, Fricker says this is not so: “This aim [of eliciting remorse] need not of

---

<sup>33</sup> Fricker.

course be present as an intention in the psychology of the communicative blamer; rather the aim is a function of the type of speech act it is, the nature of its illocutionary point.” (173) Fricker’s justification applies to what she calls a practice, a certain social form, rather than to the actions or attitudes of an individual.

How does this help? We have already seen that the mere fact that an attitude has a function or serves a purpose does not mean that I adopt it *in order to* fulfill the function or achieve that purpose: recall my trust and pride in my children. So the reactive attitudes may serve various purposes in our social lives together—perhaps they serve the functions of Fricker’s Communicative Blame: to elicit remorse, align understanding, and provide motivation. But, if these are, as I suspect, reasons of the wrong kind for these reactive attitudes, then a person could not adopt them for these reasons, nor could these reasons justify the *individual* in resenting or admiring or feeling grateful. The social function of an attitude may turn out to be the wrong kind of reason for it.<sup>34</sup>

Fricker extends the point to speech acts, which are voluntary. Perhaps it can be so extended. Though I am doubtful, I will not further explore her view here.

---

<sup>34</sup> In “Articulating an Uncompromising Forgiveness” I suggested that resentment responds, in part, to the fact that a wrong done to you, if not in some way marked as wrong, poses a kind of threat to your on-going moral status. Resentment, I said, protests the wrong done, and so marks it as a wrong. It responds to the threat. That suggestion has recently been taken up by those who see the reactive attitudes as a kind of communication. But I am also on record claiming that resentment is not a form of communication—is not a communicative act. It can seem that I am being inconsistent—protesting, one might think, is definitely a voluntary communicative act, and if resentment is a form of protest, then it should also be a communicative act.

What I would like to say is this: Resentment functions, in our social life, to mark a wrong as wrong, and so functions as a protest, but, when I resent, I am not thereby engaging in a voluntary communicative action. I do not resent in order to protest or in order to mark the wrong as wrong. Rather, I react to the threat posed to my standing or status with resentment. Although my resentment functions, in our social life, to mark the wrong as wrong, and so *functions* as a kind of protest, that is not *my* aim or purpose. (If it were, then in those cases in which I thought it best, all things considered, to keep quiet, I would, for that reason, rationally revise my resentment).

Even so, there may be some sense in which, as I resent, I am taken up, in some way, into the social form of protesting or communicating—just as, if I take pride in my children’s accomplishments, I may be in some way taken up into the project of encouraging their personal development. I think we can allow for this without saying what seems false—that pride or resentment can be adopted in order to protest or in order to encourage personal development. Clearly more should be said on this topic.

Rather, I will simply note that we need to divide questions of justification in more complex ways—in ways that Strawson anticipated. We need to heed the difference between asking whether an individual is responding as she has reason to respond, within a given system of moral expectations and reactions, and asking, instead, whether that system, itself, should be revised in some way—whether there is reason to try to change parts of it. We should heed the difference between those questions while noting they may interact—if the system is flawed, then the most ideal reaction of an individual within it may recognize that flaw in some way. Nonetheless, the two questions are distinguishable, and both must be separated from a third: the wholesale skeptical question of whether we are ought to participate in any such system, at all.

Having thus divided the question, I believe it will become clear that the first and second require careful ethical reflection. I have been at pains to emphasize that answering them will require heeding the distinction between the right and the wrong kind of reason, and so will require a more complex form of reasoning than answering questions about the permissibility or appropriateness of an ordinary action.

I will give one example of the way in which a system of moral expectations and reactions might be flawed and might be changed. I suspect that we have, at work in our culture (and in so our philosophical discussion), at least two distinguishable forms of resentment, which I have elsewhere labeled resentment-plus and resentment-minus.<sup>35</sup> Resentment-minus is a reaction to the threat posed by the disregard of a mind that matters. It carries no commitments about whether the other mind could reasonably be expected to have avoided that disregard. Resentment-plus adds this extra commitment—it commits to the claim that, if the wrongdoer had taken reasonable steps, had, e.g., paid attention and tried harder, the disrespect would have been avoided. Because of the additional commitment, resentment-plus will be rationally undermined and unjustified in cases in which resentment-minus is not. I believe that once we recognize our place in nature—once we see that we

---

<sup>35</sup> Hieronymi, "Reflection and Responsibility."; "Forgiveness, Blame, Reasons...".

are not free spirits—we will recognize that resentment-plus is often inappropriate. Often enough, given a person's personal and developmental history, a particular episode of disrespect could not have been avoided by taking steps we could reasonably expect to have been taken. Certain people are, due to their personal and developmental history, too insensitive, or too insecure, or too staked on certain ideals or institutions or ways of life, to recognize certain of their actions or attitudes as disrespectful to others. Moreover, they came to those vices in ways we could not reasonably expect them to have avoided. That is tragic, but their disrespectful action or attitude is no less disrespectful, for the tragedy. Resentment-plus is inappropriate, but resentment-minus is not.

If our previous views about our place in nature provided us, as a culture, with only resentment-plus—because we thought each free spirit could always, with a little more effort, do the right thing—then, when we learn we are mistaken, we will need to avail ourselves of an alternative. We could then recognize resentment-minus.<sup>36</sup> The system, itself, may need to be revised—though, again, this is not an individual nor even a collective decision, but more of a process. And, in that process, the most ideal response of an individual within the system might recognize the difficulty—they may, e.g., make explicit that their response does not carry the inappropriate commitment.

## 1.7 SUMMARY

Thus far, I hope to have made convincing the idea that many reactions to moral failing are non-voluntary and therefore subject to a wrong-kind-of-reasons problem and that heeding this is crucial when considering the justification of such reactions, because whether a person has adequate reason for such an attitude depends only on reasons of the right kind. In the examples considered, facts about our duties, about losses to others, and even about the nature or importance of our relationships, do not necessarily constitute the right kind of reason for adopting or changing a given attitude. And, if they are the wrong kind of reason for adopting or changing the attitude, they are

---

<sup>36</sup> I consider this possibility in "Freedom, Resentment, and the Metaphysics of Morals," (in progress).

also the wrong kind of reason for challenging it. For any non-voluntary reaction, we run the danger of either questioning its justification or attempting to justify it with reasons of the wrong kind. Moreover, determining which are the right kind of reasons is not a matter of straightforward moral reasoning. It requires careful philosophical and psychological excavation.

I then made the further observation that these non-voluntary attitudes are not a fixed part of nature, but a kind of social product, somewhat like language or symbol. They may serve certain social purposes, and they may change over time. But the purposes they serve may be the wrong kind of reason for them—it may be that an individual cannot adopt them for those reasons, as sincere or authentic forms of the attitude. And, although they may change over time, they are not directly subject to the decision-making of any individual or even collective body. Nonetheless, they are not beyond revision, and so not beyond questions of justification, as I tried to illustrate.

## **2. THE SIGNIFICANCE OF MORAL FAILURE AND REACTIONS THERETO**

But there is yet a further set of points to make. To start, rewind to 1972. Late that year, Carly Simon released a hit single, “You’re So Vain,” the refrain of which says, “You’re so vain. / You probably think this song is about you. / You’re so vain. / I’ll bet you think this song is about you, / don’t you, / don’t you?” The verses each address a wealthy womanizer from the point of view of a jilted lover. Part of the hook of the song is the apparent contradiction in its refrain: Surely the song *is* about the person to whom it is addressed, whose extravagant lifestyle is described in fine detail in its verses—how could it be otherwise? What could she mean, implying it is not about him?

The puzzle could be resolved with the thought that a different man is addressed in each verse (several candidates have been put forward, over the years). Any one man who thinks the song is about him, in particular, would then be mistaken. Or, perhaps the song is addressing an entire type, any one of which might think, wrongly, that it is specifically about him.

But another way of resolving the puzzle, one I much prefer, is with the thought that the song is not really about the person or persons to whom it is addressed, at all. It is about the singer. It is not



about the womanizer; it is about his victim, who appears in the second verse. The end of that verse is repeated after the guitar solo: “I had some dreams; / they were clouds in my coffee, / clouds in my coffee.” Like cream stirred into coffee, her dreams were dissolved into the muck of his ego. Her dreams are what he did not recognize, at the time, and will not appreciate, even now, despite her repetition of the point. He will hear her criticism, and he will correctly hear it as a criticism of him, but, vain as he is, he will miss the source of her criticism, in her mistreatment, and so miss its importance—it is criticism *of* him, but it is *about* her.<sup>37</sup>

Likewise, I would like to say, apt moral criticism is criticism of the wrongdoer, but it is about the one wronged. If you are the apt target of blame, the blame is not really about you.<sup>38</sup>

To try to come at the point in another way: moral criticism points out moral failure, but the significance of moral failure is not simply the significance of a failure of performance, even of a failure to meet some extremely important or cherished standard. Suppose you are an olympic athlete. You love your sport. You are enraptured by its beauty, or its grit, or its challenge to the limits of human ability. You spend your days and all your energy training. You have at last secured your place in the world-wide competition. When the time comes—you botch it. Your focus is off; your muscles lag; you do not fight through the pain. Your performance was below even your own personal average. Afterwards, you are devastated, filled with disappointment and self-reproach at your failure to meet the standards of human excellence that you and yours hold dear, to which you have devoted your life.

My thought is that, if this is how you experience a moral failure, or the realization that you have disregarded or disrespected another person, then you are missing something. In fact, you are

---

<sup>37</sup> On a nearby interpretation, the song is not only about the singer, but also about the other victims who appear at the end of each verse (“...they’d be your partner... clouds in my coffee... wife of a close friend...”).

<sup>38</sup> There are a handful of verbal crutches I think philosophers should try to avoid whenever possible: modifiers such as “really,” “genuinely” or “metaphysical;” metaphors such as “internal/external,” “direct/indirect,” and “point of view;” “standpoint,” or “stance;” and, finally, that particular use of italics that is meant to shift, to the reader, the burden of locating the exact sense in which a word is meant. In the last three sentences I have employed two of these crutches. At the moment, I can do no better. I hope, in later work, to retire them.

missing the most important aspect of your situation: the fact that there is another person who has been disregarded, whose own importance has been overlooked, and who therefore now needs your attention, apology, and redoubled concern. By falling into disappointment and self-reproach over your own failure, you risk once again overlooking the importance of the other person.

One might accept that the significance of moral failure lies in the one wronged, and perhaps even accept the idea that the significance of apt moral criticism lies there as well, and yet resist the idea that the same is true of the other responses to moral failure that we have been putting under the head “blame.” These responses are directed at the wrongdoer, one might think, in a different way: the wrongdoer is not merely their subject matter but also, so to speak, their target. Thus, while one might accept that simple criticism is of the wrongdoer but about the one wronged, one might insist that these other responses are about the wrongdoer in a more significant way. For example, one might think the recording and performance of “You’re So Vain” is, in fact, a clever way of taking a bit of revenge—a way for the singer to embarrass or strike back at the person who hurt her, in part by making the contradictory accusation.

Perhaps it is. But at this point we should again consider the difference between voluntary responses and non-voluntary reactions. The recording and performance of the song are voluntary actions, and, as such, they might be undertaken for an almost limitless number of reasons, including reasons of revenge, retribution, sanction, or penalty.<sup>39</sup> But, while the recording and performance of the song are voluntary, other reactions are not: resentment, indignation, distrust, or contempt are not at our discretion in the way ordinary actions are; they can be adopted only for certain reasons.

---

<sup>39</sup> Another purpose for the recording and performance of the song seems to me more likely: It is, I think, a way of re-affirming or asserting the singer’s own standing. As I see it, the song is about the singer twice over. First, as noted in the main text, it is about her dreams, feelings, and vulnerabilities as the source of the criticism, which were overlooked first in the relationship and will again be overlooked in her criticism. But, in its writing and performance, the song is also about the hard-earned confidence that enables her to showcase her vulnerability, even while anticipating that it will, again, be overshadowed by the self-concern of the person criticized—even in the ears of the audience. (The wide, self-confident grin on Carly Simon’s face, as she sings the song in the video from her 1987 concert, strikes me as a kind of confirmation. Though perhaps others will see in it the satisfaction of vengeance, I do not.)

These non-voluntary reactions are not only not well understood as sanctions or penalties, but are also, I think, not fit to be characterized, themselves, as revenge or retribution.

Resentment and indignation might seem vengeful or retaliatory simply because they are angry or in some way aggressive. But the fact that a response is angry or aggressive is not enough to secure the claim that it is retribution or vengeance. Retribution and vengeance require an intention to cause something like harm or suffering. Resentment and indignation do not.<sup>40</sup> Moreover, for a response to count as retribution, it seems to me that it must be at one's discretion in a way that would allow one to opt against it, to opt, perhaps, for mercy, and for a response to count as vengeance, it seems to me it must, in addition, be at one's discretion in a way that would allow one to plot or plan it out over time.<sup>41</sup> Thus retribution and revenge must, I think, be voluntary.<sup>42</sup>

Now, to be sure, non-voluntary reactions to moral failure, such as resentment or indignation or disappointment, can be and too often are *used* to retaliate or to exact vengeance. They are also used as sanctions or penalties. Too often, people wield their emotional reactions, in their (voluntary) speech and actions, in order either to impose suffering on or to intentionally penalize the misbehavior of coworkers, friends, and family members. I lump these regrettable activities together under the head *guilt-tripping*.

What is more, philosophers sometimes build moral theories in which, it seems, responses to moral failing are thought of as moral sanctions or penalties that we, as members of the moral community, impose on one another. In fact, I think that this is the most natural form for the merited-consequences view of responsibility to take. Thought of in this way, the blamer takes up a

---

<sup>40</sup> I argue that resentment need not involve a desire to harm in "Articulating an Uncompromising Forgiveness," pp. 359–43.

<sup>41</sup> For recent, insightful discussion of mercy, see Adam Perry, "Government Mercy," (in progress).

<sup>42</sup> Once we notice the difference between thinking that the fact of wrongdoing warrants the voluntary imposition of some hardship or suffering (whether as sanction, penalty, retribution, or revenge) and thinking that the fact of wrongdoing warrants some non-voluntary response (like resentment, indignation, or distrust), we should see that the person who insists that a non-voluntary response is not undertaken for any further purpose is not, thereby, endorsing a retributivist account of blame.

role analogous to that of parent or teacher or state or governing body of a sport, as an enforcer of moral standards—the blamer presumes some kind of authority, in imposing the sanction. Certain theories aim to explain when and why someone can rightly exercise that authority. The challenge is to avoid paternalism, manipulation, or an illicit presumption of power—to preserve autonomy within this regime.

Although I suspect that many people, in ordinary life, do think of themselves as penalizing or imposing moral sanctions in response to ordinary interpersonal wrongs—imposing negative consequences for the misbehavior of neighbors, friends, and family members—I would like to set this aside. It is, to my mind, a dubious practice, one that perhaps warrants a defensive response from its target. But this kind of moral sanctioning is not the central or most important response to ordinary interpersonal failing among adults, and reacting to a wrong done with resentment, indignation, or distrust is not, by itself, a form of sanctioning: it is not, by itself, attaching a negative consequence to a wrong for a purpose.<sup>43</sup>

Notice that we can deny that these non-voluntary reactions are forms of sanction while allowing that they may well serve the social function of, say, providing a disincentive—just as my pride functions to build my children’s self-esteem. Insofar as sanctions also function as a disincentive, the theories that characterize blame as a sanction may be close to the truth. However, once we understand these reactions as non-voluntary, we can see that, when I resent what you have done, I am not doing so *in order to* provide a disincentive—even if my resentment in fact provides a

---

<sup>43</sup> Likewise, I strongly doubt that responses of “opprobrium” or “judgment” (where “judgment” somehow goes beyond mere belief or evaluation and means something more like “personal condemnation”) are, themselves, forms of sanctioning. However, I find these responses difficult to understand—hard to theorize. I am suppressing them, for the moment, because I am not sure what to say about them—other than that I find them immensely unattractive. (I am also being purposefully silent about punishment.)

A different dynamic to note: While blamers sometimes use their emotions as a kind of sanction, from the other side, those who are charged with wrongdoing sometimes use *their* emotional reactions to the accusation (which can range from offense to distress to despair) as a way to direct attention away from the wrong, and away from the victim, and back towards themselves—to turn themselves into the victim-of-the-moment. This dynamic appears, for example, in cases of “white fragility.” See Robin DiAngelo, “White Fragility,” *International Journal of Critical Pedagogy* 3, no. 3 (2011).

disincentive and I know this.<sup>44</sup> The nature of the attitude will not allow it to be adopted for that purpose (though, again, if I feel resentment, I may then go on to use my emotion, in my voluntary speech and actions, for that purpose, or for other purposes). And so, even though my resentment may provide a disincentive, I am not imposing a sanction.

I hope I can now begin to say what seems off, or askew, about the merited-consequences conception of responsibility. The idea of “merit” can invite the thought that certain consequences are “deserved,” where the idea of “desert” is not the innocent thought that certain responses are warranted or apt or made appropriate by certain actions or attitudes, but is instead somehow caught up with the idea of retribution, penalty, or punishment.<sup>45</sup> Thus blaming can be confused with sanctioning or retribution. In response, one might clarify that the idea of “merit” is simply the more neutral claim of appropriateness or aptness (as many in the literature would insist). Even so, the tendency to characterize responses to moral failing as voluntary—as the sort of thing one can adopt or refrain from for any reason—will reinforce an already-latent tendency to see anything beyond negative evaluation as a kind of unwelcome treatment that the blamer, at their discretion, imposes on the wrongdoer. The wrongdoer, as the target of this supposedly discretionary treatment, then becomes the focus of our theoretical attention, as the passive side of the exchange, while the one wronged, the victim, who stands in need of recognition, is instead subjected to questioning, to a request for justification.

We can thus end up adopting, as theorists, the same defensive posture it is all too natural to adopt, in life, when charged with wrongdoing: when you are charged with wrongdoing, when you are the target of a blaming response, it is natural to feel something is being done to you, that you are

---

<sup>44</sup> Another example: My grief may function to elicit in you both sadness and eagerness to aid. But I do not grieve in order to make you sad or eager to aid. Cf. Fricker’s purposes of aligning understanding and motivation. In footnote 33 I noted that I may be in some way taken up into the function of building my children’s self-esteem. The same vague caveat should apply here.

<sup>45</sup> I confess that I find the use of “desert,” in the literature about responsibility and punishment, baffling.

undergoing some treatment or being punished in some way, and it is natural to object that the treatment is unjustified or unwarranted and to reach for any and all available considerations to call it into question.<sup>46</sup> Likewise, as theorists, we characterize responses to moral failing as a treatment or consequence imposed on the wrongdoer by the blamer, we question its justification, and we are tempted to allow any and all available considerations to call it into question—as we would, if it were an ordinary action.

When we do so, we show that we have not fully appreciated Strawson's thought that the reactive attitudes are *reactive*—they are, as he put it, our reactions to our perception of the quality of another's will.<sup>47</sup> These attitudes have their own logic or structure, so to speak, which is narrowly focused, by the nature of the attitude, on only certain features of the situation.

Once we appreciate that certain reactions are non-voluntary, we must conclude that they are not consequences voluntarily imposed upon the wrongdoer by the blamer. One might try to suggest that they are imposed upon the wrongdoer by the blamer in some non-voluntary way—but the language of imposition seems off, here. It seems there would be as much reason to say that the wrongdoer imposes the reactions upon the one wronged, by wronging them. Rather than think of these reactions as imposed treatment, I think it better to think of them as manifestations of the way in which people matter to other people.

In resisting the idea of imposed treatment, I am tempted to say that being the target of resentment or of indignation is more like suffering from a hangover than like being sent to your room: it is, in a sense, a natural consequence of your disrespect or disregard of others.

But, while it might be more like a hangover than a sanction, it is importantly unlike a hangover: to characterize the reaction as a purely natural consequence, like a hangover, puts the reactor in bad faith. Although these reactions are non-voluntary, they are neither involuntary nor out of one's

---

<sup>46</sup> Like judgmentalness, defensiveness has received surprisingly little philosophical attention.

<sup>47</sup> Strawson, pp. 83–84.

control—they are, in the sense earlier considered, up to the person reacting, something that person can revise and can be criticized for. Better, I think, to say they are manifestations of the way in which people matter to other people. Again, mattering, not meriting, seems to me central.

If you are found out in some wrongdoing, a blaming response is about you, insofar as it is a reaction to the perceived quality of your will, as a will that matters. But it is not—or, it is not necessarily and in the best case it is not—treatment voluntarily imposed upon you, and so not about you in that way. Nor should we think the most important thing about these reactions is how burdensome they are for the wrongdoer. In the best case, that burden is not among the blamer’s reasons and is irrelevant to their justification. What is important is what the reactions reveal. And, the important thing they reveal is not the fact that you have failed, but rather the fact that someone else was wronged. So, in another, perhaps more important sense when you are blamed, the blame is not really about you.<sup>48</sup>

---

<sup>48</sup> I am grateful to many for comments and conversation, including Rachel Achs, Facundo Alonso, Selim Berker, Sarah Buss, Jenna Donohue, Michael Hicks, Mark C. Johnson, Christine Korsgaard, Doug Kremm, Richard Moran, Gaile Pohlhaus, Darien Pollock, Ronni Sadosky, T. M. Scanlon, Angela M. Smith, my graduate seminar at UCLA in Fall 2016, the UCLA Legal Theory Workshop, the Harvard Ethics Workshop, and audiences at Harvard University, the Conference in Honor of Gary Watson, Toronto University, UC Irvine, University of Southern California, William and Mary, UT Austin, and Miami University of Ohio. Special thanks and gratitude are due to Gary Watson, for his inspiring philosophical work.

## BIBLIOGRAPHY

- Clarke, Randolph, Michael McKenna, and Angela M. Smith, eds. *The Nature of Moral Responsibility*. New York: Oxford University Press, 2015.
- DiAngelo, Robin. "White Fragility." *International Journal of Critical Pedagogy* 3, no. 3 (2011): 54–70.
- Fischer, John Martin, and Neal Tognazzini. "The Truth About Tracing." *Noûs* 43, no. 3 (2009): 531–56.
- Fricker, Miranda. "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (2016): 165–83.
- Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529–55.
- . "Believing at Will." *Canadian Journal of Philosophy, Supplementary Volume* 35 (2009): 149–87.
- . "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, no. 1 (March 2006): 45–74.
- . "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (2004): 115–48.
- . "Forgiveness, Blame, Reasons..." In *3am: magazine*, edited by Richard Marshall, 2013.
- . "Freedom, Resentment, and the Metaphysics of Morals." (in progress).
- . "Internal Reasons and the Integrity of Blame." 1996.
- . "Reflection and Responsibility." *Philosophy and Public Affairs* 42, no. 1 (2014): 3–41.
- . "Responsibility for Believing." *Synthese* 161, no. 3 (April 2008): 357–73.
- . "The Use of Reasons in Thought (and the Use of Earmarks in Arguments)." *Ethics* 124, no. 1 (October 2013): 114–27.
- . "Virtue and Its Imitation." Harvard University, 2000.
- Kavka, Gregory. "The Toxin Puzzle." *Analysis* 43 (1983): 33–36.
- McKenna, Michael. *Conversation and Responsibility*. New York: Oxford University Press, 2012.
- Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press, 2001.
- Pereboom, Derk. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press, 2014.
- Perry, Adam. "Government Mercy." (in progress).
- Rosen, Gideon. "The Alethic Conception of Moral Responsibility." Chap. 3 In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna and Angela M. Smith, 65–87. New York: Oxford University Press, 2015.
- Scanlon, T. M. "Forms and Conditions of Responsibility." In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna and Angela M. Smith. New York: Oxford University Press, 2015.
- . *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.
- Shoemaker, David. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121 (2011): 602–32.
- Smith, Angela M. "Moral Blame and Moral Protest." Chap. 2 In *Blame: Its Nature and Norms*, edited by Justin Coates and Neal Tognazzini, 27–48. New York: Oxford University Press, 2013.



- . "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. January (2005): 236–71.
- Strawson, Peter F. "Freedom and Resentment." In *Free Will*, edited by Gary Watson. Oxford: Oxford University Press, 2003.
- Wallace, R. Jay. "Dispassionate Opprobrium: On Blame and the Reactive Sentiments." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar and Samuel Freeman, 348–72. New York: Oxford University Press, 2011.
- Watson, Gary. "Standing in Judgment." In *Blame: Its Nature and Norms*, edited by Justin Coates and Neal Tognazzini. New York: Oxford University Press, 2013.
- . "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227-48.
- Williams, Bernard. "A Critique of Utilitarianism." In *Utilitarianism: For and Against*, edited by J. J. C. Smart and Bernard Williams, 75-150. Cambridge: Cambridge University Press, 1973.
- . "Internal and External Reasons." In *Moral Luck*, 101-13, 1981.
- Wolf, Susan. *Freedom within Reason*. New York: Oxford University Press, 1990.
- Zimmerman, Michael J. "Varieties of Moral Responsibility." Chap. 2 In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna and Angela M. Smith, 45–64. New York: Oxford University Press, 2015.