

Responsibility for Attitudes: Activity and Passivity in Mental Life*

Angela M. Smith

I. INTRODUCTION

I forgot a close friend's birthday last year. A few days after the fact, I realized that this important date had come and gone without my so much as sending a card or giving her a call. I was mortified. What kind of a friend could forget such a thing? Within minutes I was on the phone to her, acknowledging my fault and offering my apologies.

But what, exactly, was the nature of my fault in this case? After all, I did not consciously *choose* to forget this special day or deliberately *decide* to ignore it. I did not *intend* to hurt my friend's feelings or even *foresee* that my conduct would have this effect. I just forgot. It didn't occur to me. I failed to notice. And yet, despite the apparent involuntariness of this failure, there was no doubt in either of our minds that I was, indeed, responsible for it. Although my friend was quick to pardon my thoughtlessness and to dismiss it as trivial and unimportant, the act of pardoning itself is simply a way of renouncing certain critical responses which it is acknowledged would, in principle, be justified.

Moments such as these—which for many of us, I imagine, are more common than we would like to admit—reveal a deep tension in our ordinary thinking about the conditions of moral responsibility. If asked, most of us would probably say that choice or voluntary control is a precondition of legitimate moral assessment. And yet, as the case above was meant to illustrate, we regularly do hold ourselves and others responsible for things that do not appear to reflect a conscious choice or decision. Indeed, we quite often respond to people's spontaneous at-

* For very helpful discussion and comments on earlier drafts of this article, I am grateful to Ann Baker, Laurence Bonjour, Sarah Buss, Pamela Hieronymi, Christine Korsgaard, Rahul Kumar, Adam Leite, Richard Moran, T. M. Scanlon, William Talbott, and the readers and editors of *Ethics*. I am also grateful to the audience members in attendance for my presentation of parts of this material at Willamette University and the University of Washington.

titudes, reactions, and patterns of awareness in many of the same ways that we respond to their deliberately chosen actions—for example, with hurt feelings, gratitude, resentment, and various forms of moral and nonmoral criticism. We say things like, “You shouldn’t be so angry,” “How can you find that amusing?” or “You should have noticed her discomfort,” for instance, where these criticisms suggest that we do take people to be responsible and open to moral assessment for these seemingly involuntary responses.

I have argued elsewhere against various attempts to try to reconcile our actual moral assessments in these cases with our alleged commitment to choice or voluntariness as a precondition of moral responsibility.¹ My aim in this article is to present an alternative to what I have called the volitional view of moral responsibility, one which I think does a better job of capturing the real basis of our responsibility for our own attitudes. Since it is often claimed that the considerations that push us toward the volitional view arise out of our commonsense intuitions about the nature of activity and passivity, however, much of my argument in this article will be devoted to analyzing and rejecting this claim. I will try to show that our commonsense intuitions do not, in fact, favor a volitionalist criterion of responsibility, but a rationalist one. That is to say, I will argue that the kind of activity implied by our moral practices is not the activity of choice, but the activity of evaluative judgment. This distinction is important, because it allows us to say that what makes an attitude “ours” in the sense relevant to questions of responsibility and moral assessment is not that we have voluntarily chosen it or that we have voluntary control over it, but that it reflects our own evaluative judgments or appraisals. There are a number of different ways in which our attitudes can be said to reflect our evaluative judgments, but what is important here is that a connection to choice is not an essential condition of responsibility for these states.

My strategy in this article will be as follows. In Section II, I will explain the notion of responsibility I take to be at issue in debates over the conditions of responsibility and will briefly describe some of the different forms that a volitional view of responsibility can take. In Section III, I will consider some examples which seem both to tell against the volitional view (in all of its forms) and also to provide a strong *prima facie* case in favor of an alternative rationalist account of responsibility. In Section IV, I will describe in more detail some of the different ways in which our attitudes and reactions can be said to reflect our evaluative judgments and will explain why these connections to judgment are sufficient for ascriptions of responsibility in the most basic sense. Finally,

1. See my article “Identification and Responsibility,” in *Moral Responsibility and Ontology*, ed. Ton van den Beld (Dordrecht: Kluwer Academic, 2000), pp. 233–46.

in Section V, I will examine some lingering volitionalist objections to this account and will try to respond to some concerns about the apparently ahistorical nature of this approach to responsibility.

II. THE VOLITIONAL VIEW OF RESPONSIBILITY

Before describing some of the different versions of the volitional view of responsibility, I should say something briefly about the basic notion of responsibility I take to be at issue in this debate over responsibility for attitudes. I interpret the fundamental question of responsibility as a question about the conditions of moral attributability, that is to say, the conditions under which something can be attributed to a person in the way that is required in order for it to be a basis for moral appraisal of that person. To say that a person is responsible for something, in this sense, is only to say that she is *open* to moral appraisal on account of it (where nothing is implied about what that appraisal, if any, should be).² Although we sometimes use the term 'responsibility' in a stronger sense, to indicate an assessment of blameworthiness (as in "I hold you responsible for this mess"), this stronger use presupposes responsibility in the more basic sense of attributability. The volitional view and the rationalist view that I will defend, then, can be seen as offering competing accounts of the conditions under which something can be attributed to a person in the relevant sense.

What I have called the volitional view of responsibility is actually better understood as a cluster of distinct views which share a common assumption, namely, that choice, decision, or susceptibility to voluntary control is a necessary condition of responsibility (for attitudes as well as actions). According to all of these views, what ultimately makes an attitude attributable to a person for purposes of moral assessment is that it is connected in some way to her choices: she made choices in the past which led to the development of the attitude in question, she made a choice in the present to endorse or "identify with" it, or she has the ability to modify her attitudes through the choices she makes

2. For similar interpretations of this central idea of moral responsibility, see Bernard Berofsky, *Freedom from Necessity: The Metaphysical Basis of Responsibility* (London: Routledge & Kegan Paul, 1987), p. 45: "The core idea of moral responsibility for action A [is] the principle that A is proper to cite in a moral evaluation of the agent"; Justin Oakley, *Morality and the Emotions* (London: Routledge, 1992), p. 124: "To be responsible for something is to be open to creditworthiness or blameworthiness for it, but whether one is actually creditworthy or blameworthy for it depends also on its goodness or badness (or rightness or wrongness)"; and T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), p. 248: "To say that a person is responsible, in this sense, for a given action is only to say that it is appropriate to take it as a basis of moral appraisal of that person. Nothing is implied about what this appraisal should be—that is to say, about whether the action is praiseworthy, blameworthy, or morally indifferent."

in the future. It is the connection to past, present, or possible future choices that makes it appropriate to attribute (some) attitudes to persons for purposes of moral assessment, according to these volitionalist accounts.³

Though these views share the assumption that some connection to choice is necessary to establish responsibility for an attitude, it makes a big difference just how this connection is envisioned in these different accounts. On the “prior choice” view, what is essential for attributions of responsibility is that we be able to trace the development of an attitude to a person’s own prior choices or decisions. For example, perhaps a person made a deliberate decision to begin reading virulent neo-Nazi propaganda and then developed strongly anti-Semitic attitudes as a result. Such a person would count as responsible for her attitudes on the prior choice view, since they can be traced causally to her own past choices. Attitudes that cannot be traced in this way to a person’s own choices, however, will not count as attributable to her in the relevant sense. On the “endorsement” view, what is essential for attributions of responsibility is that the person herself endorse or identify with the attitude in question. On such views it does not matter how an attitude has come about (whether as a result of a person’s own prior choices or

3. The most common strategy in the literature for defending the claim that we are responsible for our attitudes is to argue that we are responsible in virtue of the fact that we can exert a causal influence on them through the choices we make. Some philosophers argue that we are responsible only for those attitudes whose occurrence or existence can be traced back to our own prior voluntary choices, while others think it sufficient for responsibility that we can exert (some) voluntary control over them once they appear on the mental scene. Unfortunately, it is not always clear which of these positions is being defended by those who claim that “voluntariness” or “voluntary control” is a precondition of responsibility for attitudes (see, e.g., Oakley, pp. 86–102; Larry May, *Sharing Responsibility* [Chicago: University of Chicago Press, 1992], esp. pp. 27–35; and Edward Sankowski, “Responsibility of Persons for Their Emotions,” *Canadian Journal of Philosophy* 8 [1977]: 829–40). A very different way of making choice a precondition of responsibility for attitudes is defended in the early work of Harry Frankfurt. Frankfurt rejects the view that we must be the causal initiators of our desires and other attitudes in order to be responsible for them or that we must have voluntary control over them. Instead, he argues that we become responsible for our desires and attitudes when, and only when, we make a deliberate choice to identify with them. Frankfurt presents and defends this view in a number of influential articles which were written in the seventies and eighties, including the following: “Freedom of the Will and the Concept of a Person” (1971), “Three Concepts of Free Action” (1975), “Identification and Externality” (1976), and “Identification and Wholeheartedness” (1987), all of which are reprinted in his collection *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988). These articles do not appear to represent Frankfurt’s latest views on this issue, however. For this, see his 1991 Presidential Address to the American Philosophical Association, “The Faintest Passion,” *Proceedings and Addresses of the American Philosophical Association* 66 (1992): 5–16, and some of the other articles in his more recent collection *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999).

not); all that matters is whether the person now identifies or endorses the attitude in question. If a person found herself with anti-Semitic attitudes, for example, but refused to endorse or identify with them, she would not count as responsible for them on the endorsement view (regardless of their origin). Finally, on the "voluntary control" view, what is essential for attributions of responsibility is that a person have the ability to control or modify her attitudes through her future choices or decisions. On this view, what would matter in determining whether a person is responsible for her anti-Semitic attitudes is whether she has the ability to revise or eliminate them through her own voluntary efforts. As with the endorsement view, it does not matter on the voluntary control view how the attitude came about in the first place; unlike the endorsement view, however, endorsement itself is neither necessary nor sufficient for establishing responsibility in such a case. What matters is whether the attitude is in fact susceptible to voluntary control once it appears on the mental scene.⁴

It is not my intention here to evaluate critically these different volitional accounts. Instead, my aim is to offer an alternative account of responsibility which I believe makes better sense of our ordinary intuitions about the conditions under which a person is responsible for her attitudes. I will begin my defense of this alternative account of responsibility, which I will call the "rational relations view," by considering a few examples. These examples are meant to bring out the intuitive plausibility of the rational relations view, while at the same time casting doubt upon the claim that we ordinarily take choice or voluntary control to be a precondition of legitimate moral assessment. The following section will then be devoted to further clarifying and defending the rational relations view.

III. THE SIGNIFICANCE OF SPONTANEOUS ATTITUDES

In his seminal essay "Freedom and Resentment," Peter Strawson devotes much of his argument to reminding us of a number of "commonplaces" about human relations, commonplaces which he says are "easy to forget when we are engaged in philosophy, especially in our cool, contemporary style."⁵ Of these, "the central commonplace," he claims, is "the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs

4. I am grateful to an anonymous reviewer for pressing me to say more about these different versions of volitionalism about responsibility.

5. Peter Strawson, "Freedom and Resentment," in *Perspectives on Moral Responsibility*, ed. John Martin Fischer and Mark Ravizza (Ithaca, N.Y.: Cornell University Press, 1993), pp. 45-66, p. 50.

about these attitudes and intentions.”⁶ It is worth noting that Strawson speaks here of the importance we attach not just to the deliberate choices and decisions of other human beings, but to their attitudes quite generally. We care, he insists, not only about what people intentionally choose to do, but also about the general attitudes they have and express toward us. Indeed, we often respond to people’s spontaneous attitudes, reactions, and unreflective patterns of awareness in many of the same ways that we respond to their voluntary actions. Stuart Hampshire captures well the richness and complexity of the phenomena upon which these assessments are based in the following passage, from his book *Thought and Action*:

One has before one, for reflection and comment, whether in one’s own person, or in the person of another, always a whole person, including the way he thinks and expresses his thoughts and feelings, the things that he notices and neglects, the attitudes that he adopts, the feelings that he restrains and the feelings to which he allows free play, the words that he chooses to use or that he uses unreflectingly, the gestures and physical reactions that he controls or suppresses, the plans that he makes and the sudden impulses that occur to him. All these are features of the actions and reactions of a person, upon which his own deliberation may be exercised, and upon which his own judgment, or the judgment of others, may pronounce.⁷

As Hampshire’s list makes clear, many of the features “upon which [our] own judgment, or the judgment of others, may pronounce” fall outside the scope of our immediate voluntary control. Many of these features seem neither to reflect a choice or decision nor to be susceptible to voluntary control. But if the volitional view is correct, then it would seem that our assessments should be confined to those “features of the actions and reactions of a person” that are connected in some way to her voluntary choices.

To see how implausible this is, I want to look in more detail at some of the things Hampshire mentions in his list of features that make up “the whole person”—what he notices and neglects, the thoughts and impulses that occur to him, and his spontaneous reactions. Though these things would seem to count as paradigmatic cases of “passivity” on the volitional view, we do not, in normal circumstances at least, refrain from attributing them to persons or from taking them as legitimate grounds for both moral and nonmoral assessment. Indeed, we often think it essential to a proper assessment of a person that we take into consideration these involuntary responses. Such responses provide

6. *Ibid.*, p. 48.

7. Stuart Hampshire, *Thought and Action* (London: Chatto & Windus, 1959), p. 91.

an important indication of a person's underlying moral commitments, of *who he is*, morally speaking. As a number of philosophers have pointed out, their alleged "passivity," far from undermining their attributability to persons, may actually be the strongest mark of their genuineness and sincerity.⁸ In the remainder of this section, I will look briefly at some cases which seem to support this intuition.

Noticing and Neglecting

What we do and do not notice about the world around us is not, in general, something we do by will. Of course, at any given time we can choose to concentrate or to pay careful attention to something, and this might have an effect on what and how much we notice. But it would be odd to say that we "choose (not) to notice" such and such, except when we mean by this that we intend to treat something with a particular form of (in)attention or (dis)respect (i.e., precisely when "notice" is being used to indicate a deliberate action, rather than a form of receptivity).⁹

Nevertheless, we often take what a person notices and neglects to have an enormous amount of expressive significance. Consider, for example, the following passage from George Eliot's *Scenes of Clerical Life* (1858). She is describing the dinner-table interactions of an engaged couple, Captain Wybrow and Miss Assher:

When the sweets were on the table, there was a mould of jelly just opposite Captain Wybrow, and being inclined to take some himself, he first invited Miss Assher, who coloured, and said, in rather a sharper key than usual, "Have you not learned by this time that I never take jelly?"

"Don't you?" said Captain Wybrow, whose perceptions were not acute enough for him to notice the difference of a semitone. "I should have thought you were fond of it. There was always some on the table at Farleigh, I think."

"You don't seem to take much interest in my likes and dislikes."

"I'm too much possessed by the happy thought that you like me," was the *ex officio* reply, in silvery tones.¹⁰

8. See, e.g., Robert Adams, "Involuntary Sins," *Philosophical Review* 94 (1985): 1–31, p. 10; Jonathan Bennett, "The Conscience of Huckleberry Finn," *Philosophy* 49 (1974): 123–34; Harry Frankfurt, "The Importance of What We Care About," in *The Importance of What We Care About*, pp. 80–94; and Bernard Williams, "Morality and the Emotions," in *Problems of the Self* (Cambridge: Cambridge University Press, 1973), pp. 207–29, p. 227.

9. I have in mind cases in which one says something like: "I choose not to notice his wounding remarks," where it is quite clear that what one means by this is that one chooses not to acknowledge them (and not, literally, that one is blissfully oblivious to them).

10. George Eliot, *Scenes of Clerical Life* (London: Penguin, 1973), pp. 170–71.

In her usual incisive and psychologically astute fashion, and with materials that could not be more ordinary or familiar, Eliot captures the distinctive kind of importance we commonly place on a person's unreflective patterns of awareness. Captain Wybrow's seemingly innocent efforts to be polite are interpreted by Miss Assher as a damning indication of his emotional indifference toward her. His failure to show any special awareness or appreciation of her own tastes and preferences, in a way that would distinguish her from any other lady who happened to be sitting at the table next to him, suggests to Miss Assher that she does not yet occupy a distinctive place in his overall emotional and evaluative outlook. Miss Assher's reaction, though a bit extreme perhaps, is not difficult to understand. We do expect our intimates to notice and to be sensitive to factors which pertain to our welfare in a way that we do not expect others to be; Captain Wybrow's failure to be sensitive in these ways provides at least some indication that he does not, in fact, care about her in the way a lover should.¹¹

Christine Korsgaard discusses a similar example of this kind of significance in her article "Morality as Freedom," this time involving a case of friendship. She points out that if I claim to be Charlotte's friend and to have her happiness as my end, but it "seldom or never occurs to me to do anything in particular to make Charlotte happy," then I am the legitimate target of a certain kind of criticism from her. As she puts it, "Under these circumstances, surely Charlotte would be entitled to complain that there is no real sense in which I have her happiness as my end. It would not be pertinent for me to reply that I have no direct control over what occurs to me. To find certain features of the world salient is part of our notion of what it is to have an end. To have an end is to see the world in a certain way."¹² It would "not be pertinent" for me to cite my volitional passivity in this case because Charlotte is not accusing me of a volitional failure. She is claiming that my failure to recognize and to appreciate factors which bear upon her welfare casts doubt upon the depth or genuineness of my professed concern for her.

Although the significance of what we notice and neglect comes out most clearly in the context of personal relationships, similar cases can

11. Of course, in many cases of this sort all we can really infer is a kind of social clumsiness, which need not reflect an objectionable underlying judgment. What makes this case particularly illustrative is that Captain Wybrow clearly possesses the social graces; what he lacks, and what Miss Assher objects to, is a special concern for his beloved. I am grateful to Ann Baker for helpful discussion on this point.

12. Christine Korsgaard, "Morality as Freedom," in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 159–87, p. 180.

be cited in the more narrowly "moral" domain.¹³ If I do not notice when my music is too loud, when my advice is unwelcome, or when my assistance might be helpful to others, these again can be described as "involuntary" failings; nevertheless, these failings are commonly taken to be an appropriate basis for moral criticism. These forms of moral insensitivity provide at least some indication that I do not judge your needs and interests to be important, or at least that I do not take them very seriously.

These examples suggest that we take there to be some sort of rational connection between what we notice and what we evaluate or judge to be important or significant. We might characterize the connection in the following way: if one judges some thing or person to be important or significant in some way, this should (rationally) have an influence on one's tendency to notice factors which pertain to the existence, welfare, or flourishing of that thing or person. If this is so, then the fact that a person fails to take note of such factors in certain circumstances is at least some indication that she does not accept this evaluative judgment.¹⁴ I should emphasize here, to avoid misunderstanding, that this is a quite general (and indeed necessary) principle of cognitive selectivity, and the fact that a person fails to notice certain things need not reflect any objectionable evaluative judgment on her part.¹⁵ We all have different ends, projects, and interests, and we care about different

13. I should say here that I myself am inclined to view the sorts of normative expectations implicit in the cases of love and friendship just discussed as "moral" in nature. But nothing rests upon this point, so in order to avoid unnecessary objections, I am happy to confine the term 'moral' to those expectations which do not presuppose a particular personal relationship.

14. This way of framing the conditional was suggested to me by T. M. Scanlon. I say that such failures of awareness provide "at least some indication" that the person does not accept the judgment in question, because there may be other factors at work in particular cases, e.g., the person in question may be extremely tired or under a lot of stress, which block the normal inference from what a person notices to what she cares about. If we turn from specific instances to more general patterns of awareness, however, I think that an even stronger claim can be made here, to the effect that certain patterns of awareness are themselves partially constitutive of caring about something or regarding it as important. This would make the rational connection here even tighter, and more akin to the conceptual connection I take there to be (and describe below) between specific attitudes, such as fear and resentment, and certain judgments, evaluations, or appraisals. Thanks to Richard Moran for helpful discussion on this point. For more on this issue, see Frankfurt, "The Importance of What We Care About," esp. pp. 84–85.

15. For a discussion of the role of selectivity of attention, memory, and evidence-gathering in belief formation, see Gilbert Harman, *Change of View* (Cambridge, Mass.: MIT Press, 1986); as applied to the formation of intentions and other attitudes, see Scanlon, esp. pp. 22–25, 69–70. For the role of cognitive selectivity in cases of self-deception, see W. J. Talbot, "Intentional Self-Deception in a Single Coherent Self," *Philosophy and Phenomenological Research* 55 (1995): 27–74, esp. pp. 40–43.

things in different ways and to different degrees; so, if the rational connection I have characterized obtains, it is only to be expected that we will not all notice the same things. Such variations in our patterns of awareness need not themselves imply anything positively or negatively about a person's evaluative judgments. The fact that I rarely notice the clothes people wear, for example, while my best friend can give a detailed description, down to the accessories, of what everyone she has encountered in a given day had on, seems accurately to reflect the fact that she judges fashion to be important while I do not. But I need not feel bad about having this judgment attributed to me (nor, of course, need my friend feel bad about having the contrary judgment attributed to her).¹⁶ In certain cases, moreover, the fact that a person fails to notice certain things can actually reflect on her in a morally creditworthy way. Someone who fails to notice the minor moral, intellectual, or physical flaws of others, for example, is presumably not to be criticized for this fact; her neglect of these things indicates that she does not regard them as particularly important, which is an evaluative judgment most of us would find morally admirable. Finally, it should be emphasized that the evaluative judgments I claim underlie our patterns of awareness need not be judgments that something is "objectively valuable" or "objectively important" in a way that would commit us to thinking that everyone ought to be similarly attentive to the thing in question. More often, the judgments in question should be understood as judgments of relative importance or significance: that something is important or significant to me. For example, a person can judge that a loved one's needs or interests are especially important to him without thinking that these needs and interests are "objectively" more important than anyone else's.¹⁷

What makes the failures to notice so significant in the cases of love and friendship I discussed above, then, is that we expect the participants in such relationships to place a special importance on the needs and interests of their beloveds, and these failures of awareness seem to call

16. This case is obviously underdescribed. We need to know in what sense my friend judges fashion to be "important," and in what sense I judge it not to be, in order to assess whether there are any objectionable implications involved in these patterns of awareness. For example, if my friend judges these things important to determining how she ought to treat others (e.g., with respect or disdain), that would be a problem; if I judge these things unimportant to the extent that I fail to recognize the ways in which my own forms of dress can communicate respect or disrespect to others, that would also be a problem. I take the very ease with which we can come up with descriptions under which certain patterns of awareness would be morally objectionable to bolster my claim that we normally do take such patterns of awareness to be revelatory of a person's judgments and evaluative commitments.

17. I am grateful to an editor at *Ethics* for prompting me to clarify my view on this matter.

into question the presence of that evaluative commitment. Indeed, we often take a person's patterns of awareness to provide a more accurate reflection of his true commitments in such relationships than his explicit pronouncements do, as the case of Captain Wybrow and Miss Assher brings out. Captain Wybrow's flippantly amorous response to what is, for Miss Assher, a rather serious worry about the nature of his commitment to her does nothing to cast doubt upon the inference she has drawn from his lack of awareness to his lack of special concern. To the contrary, his "ex officio" reply rather disappointingly confirms her sense of his detachment from her.

What Occurs to Us

The considerations and impulses that occur to us, in general and in our deliberations about what to do, are also not normally under our immediate voluntary control. We remember some things and forget others, particular considerations strike us as relevant to the issue at hand, and we are moved spontaneously by various desires and inclinations. All of this happens without our choice or decision.

But here again, we commonly attach a great deal of importance to these unreflective features of our mental lives. Bernard Williams provides an illustration of this fact near the end of his book *Ethics and the Limits of Philosophy*. He writes: "An effective way for actions to be ruled out is that they never come into thought at all, and this is often the best way. One does not feel easy with the man who in the course of a discussion of how to deal with political or business rivals says, 'Of course, we could have them killed, but we should lay that aside right from the beginning.' It should never have come into his hands to be laid aside."¹⁸ Why is it that we "do not feel easy" with such a person (assuming, as we must for the example to be compelling, that he was not simply joking around)? After all, he does not end up endorsing the spontaneous proposal that occurs to him. Indeed, he decisively rejects it, upon reflection. Why should it matter, then, that it seemed to him momentarily to be a viable option?

The reason we continue to feel uneasy with such a person, I submit, is that we assume that for such an appalling alternative even to "come into thought" as a serious option, this person must either not accept, or not be fully committed to, the evaluative judgment that killing others is not an acceptable way of furthering one's own interests. At the very

18. Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, Mass.: Harvard University Press, 1985), p. 185. See also Joseph Raz's discussion of such cases in his postscript to *Practical Reason and Norms*, 2d ed. (Princeton, N.J.: Princeton University Press, 1990), p. 181.

least, we may feel, he must place shockingly little value on human life, if he is willing to consider sacrificing it, even momentarily, for business or political ends. It is because we assume there to be some such connection, between the way that he conceives of his options and his underlying evaluative commitments, that we remain disturbed by this person, even after he has rejected his own proposal.¹⁹

Our spontaneous desires have a similar kind of expressive significance. Despite what Kant suggests in section 1 of the *Grounding for the Metaphysics of Morals*, the sympathetic man's spontaneous desires to help others and to spread joy around him are not merely fortunate accidents of nature, for which he deserves no moral credit; instead, they arise out of, and can only be explained against the background of, a general commitment to valuing the welfare of others and to furthering their ends, where possible. The fact that similar desires and inclinations do not occur to Kant's "cold-tempered" man can also be taken to reflect something about his underlying evaluative judgments, as Kant's own analysis makes clear. He suggests that this man's indifference to the sufferings of others is due to the fact that "as regards his own sufferings he is endowed with a special gift of patience and fortitude *and expects or even requires that others should have the same.*"²⁰ While it may be an "accident of nature" that this man is specially able to endure his own sufferings, his indifference to the sufferings of others can be taken to reflect his underlying judgment that people ought to be stoic in the face of their own misfortunes.

So here again, these examples, as well as the case of forgetting discussed at the beginning of this article, suggest that we commonly take there to be a rational connection between many of the thoughts and desires that occur to us and the evaluative judgments and commitments we accept. If we value something and judge it to be worth promoting, protecting, or honoring in some way, this should (rationally) have an influence on our unreflective patterns of thought and feeling. We commonly infer from these unreflective patterns, or from their absence, what a person really cares about and judges to be important.

19. In his essay "Moral Integrity," Peter Winch emphasizes the importance that we place upon the way in which a person conceives of the situations he confronts: "If we wish to understand the moral character of a particular man and his acts it is, often at any rate, not enough to notice that, for such and such reasons, he chooses a given course of action from among those he considers as alternatives. It may be at least as important to notice *what he considers the alternatives to be* and, what is closely connected, what are the reasons he considers relevant to deploy in deciding between them" (see "Moral Integrity," in his *Ethics and Action* [London: Routledge & Kegan Paul, 1972], pp. 171–92, p. 178).

20. Immanuel Kant, *Grounding for the Metaphysics of Morals* (1785), trans. James W. Ellington (Indianapolis: Hackett, 1983), Ak. 398–99; my emphasis.

This helps to explain why we normally feel so awful about forgetting important dates, anniversaries, or occasions: because of the normal connection between what occurs to us, on the one hand, and what we care about and judge to be of importance, on the other, we recognize that our failures in these cases can reasonably be taken to reflect a lack of appreciation for the significance of the events in question.²¹ Our reasons for acknowledging such lapses and apologizing for them, it seems, are twofold: first, we hope to reassure the other person(s) that we do still care about these things and judge them to be important (though, in reality, what we may really be doing is vowing to care more about these things in the future). Second, we hope to shore up our own (perhaps weakening) commitment to the relationships in which these events have significance.²²

I said above that we take “many” of the thoughts and desires that occur to us to reflect our underlying evaluative judgments, because it seems clear that we do not attach this kind of rational or evaluative significance to all of them. The random song lyrics, advertising slogans, and even some of the memories that run through our heads in a given day, for example, do not seem to reveal much of anything about what we value or judge to be important or significant; nor do many of the appetitive desires we experience (such as hunger, thirst, and the desire to eliminate bodily wastes).²³ We still need an explanation, of course, of what distinguishes these cases from the ones discussed above (if they are, indeed, distinct). My point here is only to draw attention to the fact that in practice we do seem to place a different kind of significance on the mental occurrence of song lyrics, on the one hand, and the serious (if only momentary) contemplation of murder, on the other;

21. It is interesting to note that our apologies in such cases are almost always accompanied by explanations of how busy or swamped we were with other things at the time (whether this is true or not), because it is terribly uncomfortable to have to admit to another person that she simply failed to make it onto our radar screen, for no apparent reason. I take it that this discomfort reflects our recognition of the normal connection between what occurs to us and what we judge to be important. Thanks to William Talbot for helpful discussion on this point.

22. I am grateful to an anonymous reviewer for this suggestion. If I am correct in claiming that our forgetting usually indicates a lack of (sufficient) care for some thing, then our apology in such a case cannot, all by itself, make us start caring in the requisite way. But it can be the first step toward this goal.

23. G. F. Schueler suggests that these desires comprise a distinct set, since each of them is connected to a specific biological need of some sort. He is reluctant to add sexual desires to this list, since it is unclear whether people or animals have a biological need for sex, strictly speaking. I think that there may be additional reasons for not grouping at least most human sexual desires along with these others, which has to do with their highly complex intentionality (see G. F. Schueler, *Desire: Its Role in Practical Reason and the Explanation of Action* [Cambridge, Mass.: MIT Press, 1995], pp. 9–10).

we place a different kind of significance on spontaneous desires to drink when thirsty, on the one hand, and spontaneous desires to help others in need, on the other. My suggestion, which I will spell out in Section IV, is that we take the latter thoughts and desires in each of these pairs to reflect a person's judgmental activity in a way that the former thoughts and desires do not.

Involuntary Reactions

Another class of responses that are not normally under our immediate voluntary control are our spontaneous emotions and other attitudinal reactions. We react with, among other things, envy, admiration, resentment, awe, amusement, regret, and gratitude to the people and events we encounter, and these reactions usually arise without any choice or decision on our part.

And yet, as with our patterns of awareness and what occurs to us, we regularly take these involuntary responses to have a great deal of expressive significance. This is true not only of the reactions of others; we also learn much about ourselves from our own spontaneous and unreflective reactions. This can, at times, be painful, as when a prejudiced reaction makes us aware of the fact that we have been harboring certain objectionable biases toward others, or a jealous reaction makes us realize that we are distrustful of someone we love. But most of the time our involuntary reactions confirm and help to sustain what we affirm as our most fundamental values and commitments.

The significance of these attitudes comes out particularly clearly in cases where our own actions have unavoidable but harmful consequences. As Williams points out in his article "Moral Luck," "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."²⁴ Regret is one way in which what we value and regard as important is revealed in situations where those commitments, for whatever reason, cannot be adequately expressed in our actions. We would feel "some doubt" about a person who failed to feel regret in these circumstances, then, because this would indicate a failure to appreciate the seriousness and

24. Bernard Williams, "Moral Luck," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), pp. 20-39, p. 28.

significance of the harm he has, however blamelessly, helped to bring about.²⁵

These examples suggest an even more direct rational connection between our spontaneous reactions and our underlying evaluative judgments and commitments. Attitudes such as contempt, jealousy, and regret seem to be partially constituted by certain kinds of evaluative judgments or appraisals.²⁶ To feel contempt toward some person, for example, involves the judgment that she has some feature or has behaved in some way which makes her unworthy of one's respect, and to feel regret involves the judgment that something of value has been lost. There seems to be a conceptual connection between having these attitudes and making, or being disposed to make, certain kinds of judgments. This helps to explain why we attach so much significance to these reactions, both in our own case and in our relations with others: unlike brute sensations, which simply assail us, our spontaneous reactions reveal, in a direct and sometimes distressing way, the underlying evaluative commitments shaping our responses to the situations in which we find ourselves.

IV. THE RATIONAL RELATIONS VIEW

The examples just discussed were meant to bring out the fact that we attach (moral) significance to a wide variety of attitudes and mental states, many of which do not arise from conscious choice or decision,

25. For other discussions of the importance of regret and agent regret, see Marcia Baron, "Remorse and Agent Regret," in *Ethical Theory: Character and Virtue*, ed. Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 13 (Notre Dame, Ind.: University of Notre Dame Press, 1988), pp. 259–81; Nancy Sherman, *Making a Necessity of Virtue* (Cambridge: Cambridge University Press, 1997), pp. 43–45; Bernard Williams, "Ethical Consistency," in *Problems of the Self*, pp. 166–86, and *Shame and Necessity* (Berkeley: University of California Press, 1993), pp. 69–70.

26. Here I am assuming a broadly cognitivist or judgmentalist view of the emotions, though I mean to be neutral on the question of just what kind of cognitive state should be seen as constitutive of emotions (e.g., beliefs, judgments, construals, takings, or "seings as"). For some recent defenses of the judgmentalist view of emotions, see Ronald de Sousa, *The Rationality of Emotion* (Cambridge, Mass.: MIT Press, 1987); Patricia Greenspan, *Emotions and Reasons: An Inquiry into Emotional Justification* (London: Routledge, 1988); William Lyons, *Emotion* (Cambridge: Cambridge University Press, 1980); and Robert Solomon, *The Passions* (Garden City, N.J.: Anchor Press/Doubleday, 1976). Cheshire Calhoun is critical of cognitivist views which make belief constitutive of emotions but defends a weaker form of cognitivism in which emotions essentially involve interpretive "seings as" in her "Cognitive Emotions," in *What Is an Emotion?* ed. Cheshire Calhoun and Robert C. Solomon (Oxford: Oxford University Press, 1984), pp. 327–42. For some important recent criticisms of judgmentalist accounts, see John Deigh, "Cognitivism in the Theory of Emotions," *Ethics* 104 (1994): 824–54; Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, Mass.: Harvard University Press, 1990), pp. 126–50; and Jennifer Robinson, "Startle," *Journal of Philosophy* 92 (1995): 53–74.

and many of which do not seem to fall under our immediate voluntary control. While it may be true that we could, over time, exert an influence on many of these patterns of awareness and response by voluntarily pursuing various projects of self-improvement, I do not think that this fact plays a role in explaining either our willingness to take these attitudes as a legitimate basis for (moral) appraisal or the nature of the appraisals we in fact make. When we praise or criticize someone for an attitude it seems we are responding to something about the content of that attitude and not to facts about its origin in a person's prior voluntary choices, or to facts about its susceptibility to influence through a person's future voluntary choices. More specifically, it seems we are responding to certain judgments of the person which we take to be implicit in that attitude, judgments for which we consider her to be directly morally answerable. If this is correct, then it is a mistake to try to account for a person's responsibility for her own attitudes in terms of their connection to her prior or future voluntary choices, because that obscures the special nature of our relation to our own attitudes: we are not merely producers of our attitudes, or even guardians over them; we are, first and foremost, inhabitants of them. They are a direct reflection of what we judge to be of value, importance, or significance. I have suggested that it is in virtue of their rational connection to our evaluative judgments that they are the kinds of states for which reasons or justifications can appropriately be requested. My aim in this section is to describe these connections in more detail and to explain why this connection to judgment is sufficient for ascriptions of responsibility in the most basic sense.

The first thing that needs to be clarified about this account is what I mean by 'evaluative judgment'. I have been using this phrase rather loosely, at times even substituting for it the phrase 'evaluative commitment'. The reason for this looseness is that I want to make clear that the judgments I am concerned with are not necessarily consciously held propositional beliefs, but rather tendencies to regard certain things as having evaluative significance.²⁷ These judgments, taken together, make up the basic evaluative framework through which we view the world. They comprise the things we care about or regard as important or

27. I am grateful to Martha Nussbaum and to an anonymous reviewer for urging me to clarify my use of the term 'judgment', which both regard as having an overly rationalistic flavor. I hope my discussion in the text puts to rest some of these concerns. I have retained the term 'judgment' rather than adopting some more inclusive category such as 'appraisal' or 'evaluation', however, because I want to make clear that these are continuing and relatively stable dispositions to respond in particular ways to particular situations, and not merely onetime assessments. It is more difficult to convey this notion of a standing commitment with the use of a term such as 'appraisal' or 'evaluation', though I certainly do want to emphasize the evaluative aspect of the judgments with which I am concerned.

significant. "Judgments" in this sense do not always arise from conscious choices or decisions, and they need not be consciously recognized by the person who holds them. Indeed, these judgments are often things we discover about ourselves through our responses to questions or to situations. For example, I may not realize, until I am faced with a choice, that I value the intellectual freedom and autonomy associated with a career in academia more highly than the economic rewards and benefits associated with a career in law. Or I may discover in some situation that I care more about being liked by others than I do about standing up for my moral principles.²⁸ Although I may never have consciously entertained these evaluative judgments, I see that they are correctly attributable to me in virtue of my own responses to the situations I confront. Of course, there is an important difference between such evaluative "discoveries" and factual discoveries of the sort one might make about the natural world (e.g., that it is raining outside). For once one realizes that one holds a certain evaluative judgment, it is open to one to determine whether one has adequate justification for that judgment and to modify it or give it up if such a justification cannot be provided.²⁹ When I say that a mental state or condition "depends upon and reflects a person's evaluative judgment," therefore, I am not claiming that the state in question must have arisen from an explicit judgment of any sort; I am claiming, rather, that the state in question can reasonably be taken to reflect an evaluative judgment on the part of the

28. Many philosophers have discussed the sense in which we can make genuine discoveries about our own values and commitments through our responses to the situations we confront. Bernard Williams, e.g., discusses the way in which coming to the conclusion that we cannot do some thing can be for us at once a decision and a discovery: "It is a decision, as being indeed the conclusion of a deliberation whether to do that thing. But it presents itself to the agent also as a discovery, because the underlying dispositions have not before been focussed through and on to that very conjunction of features. The incapacity to do this thing is an expression of those dispositions as applied to this situation *through this very deliberation*. . . . If the deliberation is sound and convincing, it is so because it is the best expression of dispositions that were there already" (see his "Moral Incapacity," in *Making Sense of Humanity* [Cambridge: Cambridge University Press, 1995], pp. 46–55, p. 52). See also Frankfurt, "The Importance of What We Care About," and "Rationality and the Unthinkable," in *The Importance of What We Care About*; and Charles Taylor, "Responsibility for Self," in *The Identities of Persons*, ed. Amelie Oskenberg Rorty (Berkeley: University of California Press, 1976), pp. 281–99, esp. pp. 289–94.

29. Moran discusses the sense in which determining what we do believe is not a purely psychological question about our mental states but also, inevitably, a deliberative question about what to believe ("Making Up the Mind: Self-Interpretation and Self-Constitution," *Ratio* 1 [1988]: 135–51). I do not take there to be any tension between this claim and the possibility of genuine discoveries about our beliefs and judgments. As I acknowledge in the text, these discoveries about our underlying judgments inevitably raise for us the deliberative question Moran highlights—namely, the question whether to go on believing or judging in this way.

person, a judgment, moreover, which it is appropriate, in principle, to ask her to defend.³⁰

But what does it mean to say that a state “reflects” a person’s evaluative judgment? The view that I am putting forward takes as its starting point the idea that some of our mental states are linked to particular judgments in such a way that, if one sincerely holds a particular evaluative judgment, then the mental state in question should (or should not) occur. The “should” in question here is the should of rationality and, therefore, marks a normative ideal which our actual attitudes may not always meet (in the same way that the “should” of *modus ponens* is a rational norm that holds regardless of whether a particular person’s beliefs conform to it). To take a simple example of the connection I have in mind: if I sincerely judge that there is nothing dangerous or threatening about spiders, I should not be fearful of them. The emotion of fear is conceptually linked to the judgment that the thing feared is in some way dangerous or threatening; therefore, my judgment that spiders are not in any way dangerous or threatening rationally entails that I should not be fearful of them. Most of the time this rational connection holds in fact: we are fearful of things we judge to be dangerous or threatening, and we are not fearful of things we judge not to be dangerous or threatening. The existence of these emotions is actually dependent upon the assessments which form their justification. If those assessments change—if, for example, we become convinced that the object of our fear is not, in fact, dangerous or threatening after all—then our fear normally extinguishes. Occasionally, however, this connection can fail: I may continue to fear spiders even though I judge them not to be dangerous or threatening in any way.³¹ When this happens, I am open to a particular kind of rational criticism—namely, to a charge of irrationality. There is a direct inconsistency between the judgments I explicitly make and the judgments entailed by the attitude I in fact hold. Fear is the kind of state that we take to be, “in principle,” sensitive to our evaluative judgments; the fact that such states are not always “in fact” sensitive to these judgments shows only that we are not always fully rational.³²

30. The central features of the account of responsibility I am defending here, in particular the focus on judgment-dependence and the appropriateness of demanding reasons or justification, are drawn from Scanlon; see esp. chap. 1, “Reasons,” and chap. 6, “Responsibility.”

31. Perhaps it is worth pointing out that lots of people quite rationally dislike spiders and wish to avoid them, not out of fear but out of disgust. These are different attitudes with quite different conceptual conditions.

32. Christine Korsgaard discusses this kind of rational entailment in her article “Skepticism about Practical Reason,” in *Creating the Kingdom of Ends*, pp. 311–34. In this article, she interprets internalism’s “motivational requirement” in the same way that I am inter-

But the charge of “irrationality” is just one, rather narrow, form of rational criticism. I mention it only because the very possibility of such criticism brings out the normative link between our attitudes and certain kinds of judgments in a particularly clear way. A much more common form of rational criticism, however, presupposes this link but does not directly concern it. This kind of criticism is applicable when a person’s attitudes depend upon and reflect evaluative judgments which themselves seem (substantively) mistaken, inappropriate, or otherwise objectionable. My fear of spiders, for example, may depend upon and reflect my actual (though not necessarily explicit) judgment that such creatures are horribly dangerous and threatening. Since such a judgment, concerning ordinary house variety arachnids at least, seems mistaken, I may be open to a mild form of criticism for the attitude which reflects it, but I do not seem to be open to the charge of “irrationality.”³³ My fear really does reflect my judgment in this case; it is just that my judgment itself seems substantively mistaken. A more serious case of this sort, from a moral point of view, would be one involving an attitude of contempt or disrespect for a certain class of people, where that attitude can be taken to reflect a judgment that such people are intellectually inferior because of their race, gender, or some other characteristic. We might say that such an attitude reflects more than one objectionable judgment, since it involves both the belief that a person’s worth is based on her intelligence and the belief that a person’s intelligence is determined by her race, gender, and so on. Here again, the person is open to rational, and in this case moral, criticism for this attitude precisely because of its rational dependence on these objectionable underlying evaluative judgments.

In the case of attitudes like fear, contempt, admiration, guilt, envy, and resentment, then, we take there to be a direct normative connection between the state in question and particular kinds of judgments or evaluative appraisals. Because of this presumed connection, we can make a direct inference from the occurrence of such states to the underlying judgments these responses reflect. Of course we may decide, in a given situation, that a person is being irrational, in which case we will say that her attitude does not reflect the full story about the evaluative judgments she accepts: although her attitude indicates that she accepts a certain judgment (e.g., that spiders are dangerous), we have independent rea-

preting the more general normative connection among our attitudes and our judgments—i.e., as setting a normative standard which will not always be met in practice (see esp. pp. 320–25).

33. Scanlon gives a helpful discussion of wider and narrower uses of the term ‘irrational’ (pp. 25–30). Compare Donald Davidson, “Paradoxes of Irrationality,” in *Philosophical Essays on Freud*, ed. Richard Wollheim and James Hopkins (Cambridge: Cambridge University Press, 1982), pp. 289–305, p. 290.

sons for thinking that she does not in fact accept it (e.g., her apparently sincere claim that she does not think that spiders are dangerous). In such a case, we are faced with an interpretive difficulty: we cannot make an unambiguous inference from her spontaneous attitude to the judgments she accepts, but we may need to qualify that inference in light of the other explicit beliefs and judgments the person holds.³⁴ What makes the charge of “irrationality” rather than “nonrationality” appropriate in such cases, however, is precisely the fact that both of these conflicting evaluations seem to be correctly attributable to her. And this also explains why irrationality of this sort is itself so deeply paradoxical. As Donald Davidson puts it, “The idea of an irrational action, belief, intention, inference, or emotion is paradoxical. For the irrational is not merely the nonrational, which lies outside the ambit of the rational; irrationality is a failure within the house of reason.”³⁵ Although I do not have a satisfactory explanation of how and why such failures occur, it seems to me that in order to preserve our sense that these are cases of genuine irrationality, we must continue to view the allegedly irrational attitudes as themselves resting in some way upon a person’s “rational” assessments (where “rational” here means simply involving our rational activity, rather than reasonable in some more substantive sense). For, as Davidson correctly points out, once we view a person’s attitude as a mere force working on her mind, “then we fail to explain, or even describe, irrationality. Blind forces are in the category of the non-rational, not the irrational.”³⁶ All of which is to say that in order to identify and attribute a mental state to a person as a belief, intention, or emotion, even an irrational one, we must locate it within a rational network of other judgments and beliefs which the person in question can be taken to accept. Irrationality occurs when the normal normative connections among these judgments break down.

In the case of other mental states, such as what we notice and what occurs to us, we take there to be a somewhat different, though still normative, connection to our underlying evaluative judgments. We assume that, if a person judges something to be important or significant, this should (rationally) have an influence on her unreflective patterns of thought and feeling concerning the thing in question. As the examples discussed in the last section were meant to bring out, if a person fails to notice or respond in the relevant ways, we take this as at least some indication that she does not accept the judgment in question.

34. I think that it would be a mistake, however, to give automatic priority to a person’s explicit judgments in such cases, given the prevalence of such things as self-deception, rationalization, wishful thinking, *akrasia*, and the like.

35. Davidson, “Paradoxes of Irrationality,” p. 289.

36. *Ibid.*, p. 299.

It is a presupposition of the view I am putting forward, therefore, that a rational creature has a coherent psychology of a certain sort, such that there are systematic rational connections between the things that happen in her psychological life and the underlying judgments and values she accepts. In order for a mental state to be attributable to a person in the sense necessary for it to be a legitimate basis of (moral) appraisal, it need not arise from conscious judgment, choice, or decision. But it must be the kind of thing that either directly reflects, or is supposed to be governed by, her evaluative judgments. This condition of moral attributability is appropriate, because moral criticism addresses a person qua rational agent and asks her to acknowledge and to defend or disavow the judgments implicit in her responses to the world around her. States that are not even in principle answerable to a person's judgment, therefore, would not be attributable to a person in the relevant sense.

This last point helps explain why nonhuman animals are not responsible for their attitudes according to the rational relations view.³⁷ Though it seems plausible to claim that some nonhuman animals experience emotions (e.g., fear, anger, and joy), it is not at all plausible to claim that these emotions are answerable to their judgments in a way that would make it appropriate to expect them to rationally justify or disavow them. In order for a creature to be responsible for an attitude, on the rational relations view, it must be the kind of state that is open, in principle, to revision or modification through that creature's own processes of rational reflection. Since human beings are arguably the only animals who have the capacity to reflect on and revise their evaluative judgments in light of rational considerations, they alone will qualify as responsible for their attitudes on this account.

This view differs substantially from the volitional view of responsibility, then, in that it rejects the idea that choice or voluntary control is the basic criterion of moral attributability for attitudes.³⁸ The problem with these volitional accounts is that they fail to sort the cases in the right way. On these views, what matters in determining whether a particular mental state properly "belongs" to a person for purposes of moral assessment has nothing to do with the nature of the mental state in question, or with its internal rational connection to a person's evaluative

37. I am grateful to an editor at *Ethics* for urging me to respond to this concern about the emotions of animals, and for suggesting this general line of response.

38. If one thinks that "choices" can, like judgments, be inexplicit, unconscious, and attributed to a person simply in virtue of her responses, then my disagreement with the volitional view would turn out to be much less significant. But I think that those who take choice to be a precondition of responsibility are attracted to this view precisely because they view choices as conscious acts, which "the person" must perform (and which therefore cannot simply be attributed to her in the way that I am saying that judgments can be).

judgments. What matters is whether that state is connected to or can be influenced by the person's choices (in the past, present, or future). This means that some of a person's spontaneous attitudes will be morally attributable to her, and some of them will not, and we may not be able to tell in a given situation which of these conditions obtains. But these views fail utterly to account for our actual practices of self- and other-assessment, which do not seem to be sensitive to these issues about choice or voluntariness. The rational relations view, by contrast, gives us a compelling rationale for our actual moral practices, explaining why it is that we attach moral significance to some of a person's mental states and not to others. According to this view, it is those mental states which we regard as normatively connected to a person's evaluative judgments that we take to be attributable to her for purposes of moral appraisal.

But what mental states, if any, get ruled out by the rational relations view? Nonintentional mental states, such as physical pains, sensations, and physiological conditions such as hunger or thirst, do not seem to be normatively connected to our judgment in the relevant sense. That is to say, we do not expect these states to be rationally sensitive to our evaluative judgments or our wider cognitive and evaluative commitments. I may think that headaches are generally a "bad thing," for example, but I do not expect that evaluative judgment to have any direct effect on when and how often they occur. For this reason, it would make no sense to ask me to defend or justify my having a headache, nor would it make sense to call a headache "unreasonable" or "inappropriate."³⁹ We are essentially "passive" with respect to these states, then, because they are not the kinds of states that either directly reflect or are supposed to be governed by our underlying evaluative judgments.⁴⁰

One might object here that many of these physical states do, in fact, seem to be directly connected to our evaluative judgments. The nausea that I feel before having to speak in public, for example, seems to be a direct result of my evaluative judgment that such public performances are both important and also fraught with opportunities for

39. It would make sense to ask me to defend or justify the voluntary actions which may have brought that condition about, however, since we do take our intentions to be sensitive to our evaluative judgments. The volitional view essentially treats our desires, emotions, and other attitudes like headaches: we can be asked to defend or justify the voluntary actions which may have brought these "conditions" about, but we cannot be asked to justify the attitudes themselves.

40. Moran defends a similar account of our "essential passivity" with respect to our sensations in his "Frankfurt on Identification: Ambiguities of Activity in Mental Life," in *The Contours of Agency: Essays on Themes from Harry Frankfurt*, ed. Sarah Buss and Lee Overton (Cambridge, Mass.: MIT Press, 2002), pp. 189–217.

failure.⁴¹ The butterflies that I feel in my stomach before boarding a roller-coaster also seem to be a direct result of my evaluative judgment that such a ride is scary and somewhat dangerous. Does it follow on my account, then, that these physical states are also attributable to me for purposes of moral assessment?

These are interesting and difficult cases. I think that it is clear that some of our physical reactions can and do serve as “moral indicators” of our evaluative judgments. But I also think that it is important to recognize how our responsibility for such physical states differs from the kind of responsibility we have for our intentional mental states, such as our desires, emotions, and other attitudes. The relation between a person’s physical states and her evaluative judgments is purely causal. Physical states are not themselves (internally) judgment sensitive, though they can perhaps be caused or brought about by judgment-sensitive states. For that reason, the kind of responsibility we have for them is essentially indirect: it flows from the responsibility we have for the evaluative judgments which constitute their causal triggers. If in some circumstance this causal connection fails to hold, moreover, we are not open to rational criticism for this fact (since there can be no rational demand placed on us that our physical states conform to our judgments). In the case of attitudes like shame, jealousy, fear, and admiration, by contrast, the evaluative judgments are themselves partially constitutive of the attitudes in question. Our attitudes are not merely the causal effects of our judgments (in the way that some of our physical reactions may be causal effects of our judgments). They are, rather, active states, in the sense that they essentially involve our judgmental activity. So our responsibility in these cases is direct and normative: we have a responsibility for the internal rational justification of the attitude itself. This distinct kind of responsibility is shown, again, by the fact that it would make no sense to ask a person to defend or justify her feeling of nausea (as opposed to simply explaining its cause), though it does make perfect sense to ask her to defend or justify her shame, jealousy, fear, or admiration.

Now, part of the reason that it would make no sense to demand justification in the case of sensations and other nonintentional mental states is precisely because they are not directed upon an object or state of affairs, and hence the idea of “getting it wrong” or “being justified” in the experiencing of the state does not really have application. Directedness upon an object, or intentionality, then, seems to be a necessary condition of direct responsibility in the sense I am trying to capture. Is it also a sufficient condition? Does the fact that a mental

41. I am grateful to an anonymous reviewer for raising this important objection, and for urging me to clarify my position on this matter.

state has intentional content imply that it is normatively connected to our judgment in a way that makes demands for reasons or justifications, in principle, appropriate? At first blush, it seems not. Visual perceptions have intentional content, and yet we do not normally expect these experiential states (as distinct from the perceptual beliefs formed on the basis of them) to be sensitive to our rational judgment. Even if I rationally judge that the fact that a stick looks bent in water does not count in favor of the belief that it really is bent, I do not expect this judgment to have an effect on my visual experience. The stick will still *look* bent to me, and I am not open to rational criticism for this fact (although I certainly am open to rational criticism if I continue to form false beliefs on the basis of this perceptual experience, which I know to be nonveridical). Here again, I am essentially passive in relation to the contents of my visual experience.

The case of perception is complicated, however, for the following reason. As my discussion of noticing and neglecting above was meant to bring out, we do take people to be open to criticism for failing to notice morally (or interpersonally) salient features of the situations in which they act. A person who systematically fails to notice features of situations which bear on the welfare of others, for example, does seem to be guilty of a normative failure of some kind. Philosophers sometimes refer to this as a failure of “moral perception,” which itself suggests that our perceptions are attributable to us for purposes of moral assessment.⁴² But these are usually cases of what might be called “seeing under an aspect,” where what is in question is not a person’s visual perception *per se*, but the significance (or lack of significance) that she attaches to what she perceives. A morally insensitive person may, in a literal perceptual sense, “see” exactly the same thing as a morally sensitive person—for example, that a person is standing on a crowded subway with two very full grocery bags. What differs is that the morally sensitive person sees this person *as* uncomfortable and in need of a place to sit down, while the morally insensitive person does not.⁴³ When I say that

42. For discussions of moral perception, see Lawrence Blum, *Moral Perception and Particularity* (Cambridge: Cambridge University Press, 1994); Barbara Herman, “The Practice of Moral Judgment,” in *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1993), pp. 73–93; John McDowell, “Are Moral Reasons Hypothetical Imperatives?” *Proceedings of the Aristotelian Society* 52 (1978): 13–29, “Virtue and Reason,” *Monist* 62 (1979): 331–50, and “Two Sorts of Naturalism,” in *Virtues and Reasons: Philippa Foot and Moral Theory*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Oxford University Press, 1995), pp. 149–79; Iris Murdoch, “The Idea of Perfection,” in *Sovereignty of Good* (London: Routledge, 1970), pp. 1–45, esp. pp. 36–41; and David Wiggins, “Deliberation and Practical Reason,” *Proceedings of the Aristotelian Society* 76 (1975–76): 29–51, esp. pp. 43–49.

43. I take this example from Blum, “Moral Perception and Particularity,” in *Moral Perception and Particularity*, pp. 32–33.

visual perception is not in general judgment dependent, then, I mean to be making the fairly limited point that the contents of our visual experience are not normally sensitive to our evaluative judgments. But even this may be too strong of a claim: it may be that moral categories of this sort do, in fact, shape what we see (and not merely the significance we attach to what we see), in which case these more robust "moral perceptions" will count as morally attributable to us on the view I am advocating.

Are there other intentional mental states that do not meet the condition of judgment dependence? I suggested in the last section that some of the random thoughts and mental images that occur to us may not bear rational relations to our judgments and evaluative commitments. When a certain smell brings to mind an image from one's childhood, or certain slogans or song lyrics keep running through one's consciousness, these occurrences do not seem to reflect or depend in any way upon one's rational or judgmental activity. That is to say, these states do not seem to involve one's "taking" or "construing" things to be a certain way or to have a certain significance, and therefore we cannot draw any inferences from the occurrence of these states to the evaluative judgments one accepts. What distinguishes thoughts or mental images of this sort from the kinds of patterns of awareness I focused on in the last section is that the latter do seem to involve rational or evaluative "takings" (or "nontakings") of the relevant sort: the man who sees, if only momentarily, the advantages of having his political or business rivals dead as a reason that counts in favor of killing them; the "friend" who fails to see the fact that something would make her friend happy as a reason that counts in favor of doing it; the sympathetic man who sees another person's need as a reason that counts in favor of helping her; and so on. These sorts of thoughts and "noticings," unlike the random thoughts or phrases that sometimes run through our heads, do seem to bear rational relations to our underlying judgments and evaluative commitments. This is shown in part by the fact that it makes sense to ask a person to defend or justify her evaluative construals, in a way that it would not make sense to ask a person to defend or justify the occurrence of a mere mental image or a song lyric.

There are difficult issues in this area, however, and it may not always be clear whether certain thoughts and mental images should be taken to bear rational relations to a person's underlying judgments. What should we say of the person who is repeatedly struck by violent or pornographic images, for example, or who has a continuous stream of nasty epithets running through his consciousness? Are these also just "random thoughts," which reveal nothing about his values and underlying commitments? Insofar as there is a pattern here, I think that we would want to say that the occurrence of such thoughts does reflect an underlying

preoccupation with violence or sex and, hence, a judgment that these things have a certain kind of value or importance.⁴⁴ And, of course, dwelling on such thoughts, because one finds them pleasurable or gratifying in some way, clearly involves “takings” of the sort I have suggested distinguish mere thoughts from attitudes which can be taken to express our evaluative commitments.⁴⁵ The more random and isolated these occurrences are, however, the less plausible it may seem to say that they reflect a person’s judgments and evaluative commitments.

In addition to perceptions, stray thoughts, and mental images, it seems that an attitude “implanted” by a mad scientist, or one induced through posthypnotic suggestion, would also fail to meet the rational relations condition I have described. As in the case of many visual perceptions, we do not expect (at least initially) such intentional mental states to reflect or to be governed by the agent’s own evaluative judgment.⁴⁶ Since these attitudes are, by hypothesis, detached from a person’s own rational assessment, it would be inappropriate to demand that she defend them, or to take them as a basis of rational or moral criticism. They do not really “belong” to her in a way that would make

44. But just what kind of value or importance will depend upon the specifics of the case. A person may have such recurring thoughts not because she places a positive value on violence or sex, but, e.g., because she lives in constant fear of being violently attacked (perhaps as a result of being victimized in the past, or because she lives in a war zone). Indeed, one can even imagine cases in which a person is repeatedly plagued by such thoughts precisely because she regards them as sinful and forbidden (i.e., her anxiety not to think of them leads her to think of them all the more). So one cannot infer from the mere recurrence of a thought or image that the person in question places a positive evaluation on the thing in question. But it does seem reasonable to say that recurring thoughts of this sort do indicate that the person attaches evaluative significance of some kind to the thing in question. I am very grateful to Jeff McMahan for prompting me to clarify my claim here.

45. It may well be the case that a person who “dwells” on such thoughts is open to more serious moral criticism than someone who is constantly beset by such thoughts but finds them unwelcome and tries to eliminate them. But I think that this difference should be registered at the level of moral assessment rather than at the level of attributability. The latter person’s thoughts do not “belong” to him any less, though his attempts to eliminate them reflect on him in a morally creditworthy way. For an interesting discussion of some of these issues with specific reference to the clinical literature on psychopathology, see George Graham and G. Lynn Stephens, “Mind and Mine,” in *Philosophical Psychopathology*, ed. George Graham and G. Lynn Stephens (Cambridge, Mass.: MIT Press, 1994), pp. 91–109.

46. I say “at least initially,” because if an agent becomes aware of these attitudes and shows no tendency to revise or reject them in light of her other beliefs and commitments, we may eventually conclude that these attitudes do accurately reflect her judgment. For a good discussion of the difficulties surrounding the idea of nonrationally induced beliefs, see Daniel Dennett, “Mechanism and Responsibility,” in *Essays on Freedom of Action*, ed. Ted Honderich (London: Routledge & Kegan Paul, 1973), pp. 159–84, pp. 177–79; see also Scanlon, p. 278.

it possible to draw an inference about the evaluative judgments she accepts.

One might well ask what it is in these “implantation” cases that makes the judgments in question “not the agent’s own.” And one might also wonder how these cases differ from cases of irrationality, in which the judgments reflected in a person’s attitudes run counter to some of the other judgments she claims sincerely to accept. What differentiates implanted attitudes from these others, in my view, is that such attitudes are not based upon the agent’s own evaluative appraisal of her situation and surroundings but are induced in a way that bypasses her rational capacities altogether. They do not, therefore, reflect the agent’s own evaluative judgments and commitments. Even in cases of irrationality, I would argue, the irrational attitude reflects a genuine tendency on the part of the agent to see certain considerations as counting in favor of certain responses. Implanted attitudes, by contrast, tell us nothing about the agent’s rational or evaluative dispositions. I see no other way of giving content to the expression “the agent’s own” here, however, except in a way which makes reference to the very network of beliefs and attitudes which I am suggesting ground our attributions of responsibility. This may seem to make this account of responsibility objectionably circular and uninformative, but I think that it is what makes it specially plausible and attractive: the seeming circularity is itself a reflection of what many philosophers have referred to as the “holistic” character of the mental.⁴⁷ A reasonable account of the conditions of responsibility should preserve our sense of the rational interrelations among our attitudes, rather than treating these things as isolated entities, each of which must meet some further condition before it can be considered attributable to a person for purposes of moral assessment.

According to my analysis, therefore, a mental state is attributable to a person in the way that is required in order for it to be a basis for moral appraisal if that state is rationally connected in one of the relevant ways to her underlying evaluative judgment. This rules out most sensations and visual perceptions, random thoughts and mental images, appetitive desires, and “implanted” attitudes, since we do not expect these states to reflect or to be governed by our judgment. What gets ruled in, however, are ordinary cases of belief, intention, most desires,

47. For a strong statement and defense of this idea, see Donald Davidson, “Mental Events,” in *Essays on Actions and Events* (Oxford: Clarendon, 1980), pp. 207–27. For similar claims and discussion, see Daniel Dennett, “True Believers: The Intentional Strategy and Why It Works,” pp. 150–67, “Making Sense of Ourselves,” pp. 184–98, both in *Mind and Cognition*, ed. William G. Lycan (Oxford: Blackwell, 1990), and “Mechanism and Responsibility,” esp. p. 180; Stuart Hampshire, “Disposition and Memory,” in Wollheim and Hopkins, eds., pp. 75–91; and Richard Moran, “Interpretation Theory and the First Person,” *Philosophical Quarterly* 44 (1994): 154–73.

fear, indignation, admiration, and guilt, among others, as well as our moral perceptions and various patterns of unreflective thought and feeling which we take to be sensitive to and expressive of our underlying values and commitments. Since explicit choice or voluntary control is not necessary for these rational connections to judgment to obtain, this account implies that we can be responsible for our spontaneous attitudes and reactions no less than for our explicit practical and theoretical conclusions.

We can see more clearly, now, the sense in which we must be “active” with respect to a mental state in order to be responsible for it. According to the volitional view, the dividing line between activity and passivity goes through choice or decision: I am active, and responsible, for anything that (i) is causally related to my prior choices, (ii) I have chosen to identify with, or (iii) is susceptible to my (perhaps long-term) voluntary control. According to the view I am defending, by contrast, the dividing line between activity and passivity goes through judgment: I am active, and responsible, for anything that falls within the scope of evaluative judgment (i.e., anything that is, or should be, sensitive to my evaluative judgments and commitments). Our deliberate choices certainly fall within this scope, but, as I have tried to show, our desires, emotions, and other attitudes do as well. Nothing much turns on our use of the terms ‘active’ and ‘passive’ here, however. For those who associate “activity” closely with deliberation and choice, my view implies that we can be responsible for things with respect to which we are (volitionally) passive. For those who associate “activity” closely with responsibility, my view implies that we must have a broader understanding of activity than the one implicit in the volitional view.

We should accept this account of responsibility, I maintain, for two reasons: first, because it provides a compelling rationale for our everyday practices of both moral and nonmoral assessment; and second, because it gives us a satisfying account of the boundaries of the moral self. One of the central aims of this article has been to show that we express our moral agency and activity not only in our explicit choices and decisions, but also in what we unreflectively think, feel, desire, and notice. Theories which make choice or voluntary control a precondition of moral responsibility, in my view, leave us with an impoverished conception of moral personhood.

If the rational relations view provides the correct account of the conditions of moral attributability, however, then why is it so widely assumed, by both philosophers and nonphilosophers alike, that choice or voluntary control is a precondition of legitimate moral appraisal? An adequate defense of the rational relations view should also provide a diagnosis of these opposing volitionalist tendencies in our thought. In the next section, I will try to offer such a diagnosis, in the course of

spelling out in more detail why it is that judgment dependence, rather than choice or voluntariness, is sufficient for ascriptions of responsibility in the most basic sense.

V. RATIONAL ACTIVITY, VOLITIONAL CONTROL, AND CHOICE

If I am correct in claiming that there are normative connections linking the vast majority of our desires, attitudes, and patterns of awareness to our evaluative judgments, it should be clear why, in practice, we take them to be attributable to persons for purposes of moral assessment. It is not because we think that the occurrence of these attitudes is a consequence of prior choices or actions on the part of the person we are assessing, because we think that the person has made a conscious choice to "identify with" these desires and reactions, or because we think that the person can exert voluntary control over them. Rather, it is because we take these attitudes, like our explicit choices and decisions, to reflect our own judgmental activity. As I have said a number of times, we need not assume that these judgments are, or ever were, explicit: we may never have consciously entertained the judgments in question, and we may even be surprised by our own reactions in certain circumstances, precisely because they reveal rational or evaluative commitments which we were hitherto unaware of holding. But even in these cases, we do not normally question the fact that these reactions do reflect our own judgments. This is shown by the fact that, when these reactions seem inappropriate, we are embarrassed by them and feel compelled to correct the judgments they reflect.⁴⁸

But why, then, is it so commonly assumed that choice or voluntary control is a precondition of responsibility for attitudes? I believe that this assumption can be traced to two distinct sources. The first source is a very common but, I think, misleading analogy between bodily movements and psychological states such as desires, emotions, and other attitudes.⁴⁹ In writing about questions of responsibility, philosophers often take responsibility for physical acts as the paradigm case and then attempt to extend this analysis to the case of responsibility for attitudes. One of the things that seems to distinguish bodily movements we "make," from those that merely "happen" in or to us, is that the former are under our voluntary control in a way that the latter are not. We can normally choose whether, when, and in what way to move our bodies, and this is important to the kind of responsibility we take ourselves to

48. When the reactions in question seem morally objectionable, we are not only embarrassed by them but often feel guilty. This is typically true of attitudes like *schadenfreude*.

49. I examine this analogy in more detail in "Identification and Responsibility," pp. 239–42.

have for these movements. Physical movements over which we lack such voluntary control (e.g., seizures and spasms) are not normally considered to be attributable to us for purposes of moral assessment, precisely because we lack voluntary control over them. But our attitudes are also not normally under our voluntary control. We cannot normally choose whether, when, and in what way to have a desire, emotion, or belief. So if we extend our analysis of responsibility for physical acts to the case of responsibility for attitudes, it seems that they should not be considered attributable to us for purposes of moral assessment either. And indeed, many philosophers have pointed to our lack of voluntary control over our desires, emotions, and other attitudes as a consideration which tells against our (direct) responsibility for them.⁵⁰

But this extension of the condition of voluntary control to the psychological realm is a mistake. Insofar as we take our beliefs and other attitudes to be dependent on our rational judgments, we do not think that we are free to control or modify them in any way we choose (as we are free, within certain physical limits, to move our bodies in any way we choose). The reason we do not have volitional control over these states is precisely because they are (normally) dependent upon our rational or evaluative assessments, and therefore they will not change unless and until we become convinced that those assessments are in some way mistaken. This means that the particular kind of responsibility we have for our attitudes should not be understood in terms of the ability to produce, retain, or eliminate such states "at will." Insofar as the assumption that choice is a precondition of responsibility for attitudes rests upon this mistaken conflation of voluntaristic control and rational control, it should be rejected outright.

The assumption that choice is a precondition of responsibility for attitudes can also be traced, I think, to another source. This source is the familiar and appealing Kantian notion of moral personhood, according to which one's true or proper self is to be identified with one's practical will and not with anything that merely happens in one's mental life. As Kant puts it in section 3 of the *Grounding for the Metaphysics of Morals*: "Incitements from inclinations and impulses (and hence from the whole nature of the world of sense) cannot impair the laws of [a person's] willing insofar as he is intelligence. Indeed *he does not even hold himself responsible for such inclinations and impulses or ascribe them to his proper self*, i.e., his will, although he does ascribe to his will any indulgence which he might extend to them if he allowed them any

50. See, e.g., R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1996), pp. 131–32. For other explicit statements of this view, see Henry Sidgwick, *The Methods of Ethics* (Indianapolis: Hackett, 1981), pp. 60–61; and Richard Taylor, *Good and Evil* (New York: Macmillan, 1970), pp. 241–52.

influence on his maxims to the detriment of the rational laws of his will."⁵¹ On Kant's view, I am not responsible for my desires and inclinations—or indeed, for anything I merely experience; I am only responsible for what I actively do, for the choices I make. Though I am responsible for how I choose to respond to the “incitements” of my sensuous nature, the incitements themselves are not attributable to me as a person.⁵² What is appealing about this account is that it allows us to acknowledge the ways in which our biological nature, our culture, and the conditions in which we were raised can serve to shape our patterns of evaluation and judgment. Since many of these patterns were acquired prior to the “age of reason,” it may seem inappropriate to take these things as a basis for moral appraisal of a person. We should focus instead on how the adult person chooses to deal with these propensities, rather than on the propensities themselves, which may be due to early attachments and experiences over which the person had no rational or volitional control. This perspective infuses much of the self-help literature and approaches to therapy in vogue today, which insist upon drawing a sharp distinction between how one “feels” and how one “acts” (where only the latter is taken to be something for which one bears responsibility).

I think that this constitutes a powerful challenge to the rational relations view, a challenge which deserves a much fuller response than I will be able to give here. I do want to say something briefly, however, about the general line I would take in trying to respond to this concern.⁵³

First of all, I think that it is very important not to conflate claims about responsibility and claims about blameworthiness. As I mentioned at the beginning of Section II, the rational relations view is an account of the conditions of responsibility in the sense of moral attributability, that is, the conditions under which something can be attributed to a person in the way that is required in order for it to be a basis for moral appraisal of that person. Merely claiming that a person is responsible for something, therefore, does not by itself settle the question of what appraisal, if any, should be made of the person on the basis of it. To

51. Kant, Ak. 457–58; my emphasis.

52. I am grateful to Nussbaum for reminding me that the Greek Stoics also made this the basis for their account of responsibility for emotional attitudes. Our responsibility for our emotions, according to these thinkers, stems from the fact that we have the power to assent or not to assent to the way in which the world strikes us. But the world's appearing to us in one way or another is not something for which we can be held accountable. As should be clear by now, I believe that the world's “appearing” to us in certain ways already involves our judgmental activity and, therefore, is something for which we can appropriately be held accountable.

53. Thanks to Nussbaum and to an anonymous reviewer for pressing me to address this issue here.

say that a person is responsible for an attitude, on the rational relations view, is simply to say that that attitude is, or should be, sensitive to her evaluative judgments and that she therefore can properly be asked to defend or justify it.

My claim, then, is that what matters in determining whether an attitude is morally attributable to a person is whether it can reasonably be taken to reflect that person's evaluative judgment(s). The fact that a person's evaluative judgments are usually shaped in various ways by her early attachments and environment does nothing to undermine the claim that they are still genuinely *her judgments*—that is, they are genuine tendencies on the part of the person to place evaluative significance on certain features of the situations she confronts. And it is this feature of judgment-dependence that makes it appropriate, in principle, to ask a person to defend or disavow her attitudes. When it comes to determining attributability, then, the question is not how a person's evaluative judgments were formed but simply whether the attitude in question can legitimately be taken to reflect these judgments.

But facts about the conditions under which a person's evaluative judgments were formed may be quite relevant when it comes to questions of moral assessment. Consider a person, Abigail, who is raised in a family or community that is deeply racist, say, or religiously intolerant. It would not be at all surprising for Abigail to develop evaluative tendencies and corresponding attitudes in line with those she sees operative in her family or surrounding community. As an adult, her attitudes may continue to reflect the vicious evaluative judgments thus formed in her childhood. Now contrast this person to another, Bert, who was raised in a loving and tolerant home and community, but who later in life reflectively comes to adopt racist and intolerant values. Bert's attitudes, like Abigail's, now reflect these vicious evaluative judgments, though unlike Abigail he formed these evaluative dispositions after he reached the age of rational maturity. What should we say about these two people?

First of all, I think that we can acknowledge that Bert is, while Abigail is not (or is not solely), responsible for becoming a racist or intolerant person. The seeds of Abigail's racism-intolerance were planted well before she was capable of rationally reflecting upon these evaluative commitments, while Bert's racism-intolerance is the result of his own mature reflective endorsement. So if we are asking whether each of these people is responsible for becoming racist or intolerant, our verdict might well be mixed. But this question of responsibility (namely, the responsibility one has for becoming a certain kind of person) must be distinguished from the question of one's responsibility for the attitudes one in fact holds. In order to regard an attitude as attributable to a person, and as a legitimate basis for moral appraisal, we need not also claim that a person is responsible for becoming the

sort of person who holds such an attitude.⁵⁴ That is a separate question according to the view I am putting forward. What matters, according to the rational relations view, is that the attitude is in principle dependent upon and sensitive to the person's evaluative judgments. If a person continues to hold the objectionable attitude even after she has reached rational maturity, it is reasonable to attribute that attitude to her and to ask her to defend the judgments it reflects. It is worth noting here that if a person responded to such a demand by saying, "I am not responsible for my attitude—I was just raised this way," we would not feel compelled to withdraw our criticism. Citing the origin of one's attitude is irrelevant when what is in question is its justification.⁵⁵

Our understanding of the circumstances in which a person's evaluative tendencies were formed may, however, have a very important influence on the kind or degree of moral criticism we think it appropriate to make. We can appreciate how difficult it might be for Abigail to come to recognize the viciousness of her own evaluative judgments, given their early entrenchment in her psyche, and also how difficult it might be for her to modify these judgments once their viciousness is recognized. For this reason, we are likely to be less critical of Abigail than we are of Bert, who adopted his racist-intolerant commitments in a fully reflective way (after being exposed to the morally appropriate values of tolerance and inclusiveness). But saying that we should be sensitive to such facts when it comes to moral criticism and assessment is quite compatible with claiming that both Abigail and Bert are fully responsible for their attitudes and for the judgments they reflect. We need not claim that Abigail lacks responsibility for her attitudes in order to distinguish her case from that of Bert. Indeed, to deny someone responsibility in a case of this sort strikes me as somewhat patronizing, insofar as it suggests that the person, because of her upbringing, is

54. For a good discussion of this issue, to which I am here indebted, see Scanlon, pp. 283–85. Others have defended similar claims; see, e.g., Michele Moody-Adams, "On the Old Saw That Character Is Destiny," in *Identity, Character, and Morality*, ed. Owen Flanagan and Amelie Rorty (Cambridge, Mass.: MIT Press, 1990), pp. 111–31, esp. pp. 121–26; and Harry Frankfurt, "Identification and Wholeheartedness," pp. 171–72. For the contrary view, with specific reference to the case of unfortunate formative circumstances, see Randolph Clarke, "Free Will and the Conditions of Moral Responsibility," *Philosophical Studies* 66 (1992): 53–72, esp. pp. 62–63; and Wallace, esp. pp. 163–66, 231–35.

55. Unless, of course, the origin is intrinsically relevant to its justification, as when we cite another person's testimony as the basis for one of our own beliefs. But in this case we are appealing to the origin of our belief as a way of justifying it, rather than as a way of exempting ourselves from responsibility. If the person with the vicious attitude, likewise, were to appeal to his background as a justification for his attitude, then again we would have no reason to withdraw our criticism. Indeed, in this case we would have two grounds for criticism, in that the attitude itself is morally objectionable and the reasons for which the person holds it are bad ones.

literally incapable of appreciating and responding to rational criticism directed at her evaluative judgments.

This point is worth emphasizing, since the literature on moral responsibility is largely driven by worries about the “fairness” of holding people responsible for their actions and attitudes.⁵⁶ Putting the question in this way suggests that being held responsible is a kind of burden, something we should all like to escape, if possible. But being held responsible is as much a privilege as it is a burden. It signals that we are a full participant in the moral community, someone who is capable of regulating her own attitudes and conduct, and who therefore can legitimately be called to account for them. As Korsgaard notes, “holding one another responsible is the distinctive element in the relation of adult human beings. To hold someone responsible is to regard her as a *person*—that is to say, as a free and equal person, capable of acting both rationally and morally.”⁵⁷ To deny someone this status, then—even with regard to just one or a few attitudes—is a serious matter. It suggests that she is a passive victim of forces beyond her control, someone to be pitied and treated, perhaps, but not to be reasoned with or regarded as an appropriate participant in practices of interpersonal justification. For anyone who has had the unpleasant experience of having her emotions or reactions dismissed in this way, it should be clear that being denied responsibility for one’s attitudes has its costs. Such denials can be deeply patronizing and disrespectful, and we should not be too eager to resort to them, either in our own case or in our treatment of others.

I conclude from this that the rational relations view, far from having trouble acknowledging the historical dimension of attitude formation, provides the most compelling account of how we ought to recognize this fact in our moral practices. The fundamental question, on this view, is whether it is appropriate, in principle, to ask a person to give a rational defense of her attitudes. Even when we see that an attitude rests on evaluative judgments that were formed very early in life, it does not seem inappropriate to expect a mature rational agent to defend (or disavow) these judgments. But we can also appreciate how difficult it can be for a person to confront and modify these deeply entrenched evaluative tendencies when it comes to questions of moral criticism and assessment. Those of us who were lucky enough to be raised in homes and communities that fostered the development of morally appropriate values and commitments should not be too self-righteous in our con-

56. Wallace, e.g., says that the debate over moral responsibility should be understood in “normative terms,” which he characterizes as follows: “What are the conditions that make it morally fair for us to adopt the stance of holding people responsible?” (p. 5).

57. Christine Korsgaard, “Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations,” in *Creating the Kingdom of Ends*, pp. 188–221, p. 189.

demnation of those whose formative circumstances and resulting evaluative commitments were less fortunate. But we should also not make the patronizing mistake of assuming that a person who has had a bad upbringing is not really responsible for her attitudes and for the judgments they reflect. It is much more respectful to criticize a person for an objectionable attitude than it is to dismiss it as the inevitable product of her upbringing, her culture, or her hormones.

VI. CONCLUSION

Let me conclude, then, by giving a brief summary of the central claims I have tried to establish in this article and a sketch of some of the questions that still need to be addressed in a full defense of the rational relations view. My basic argument for the rational relations view might be summarized as follows:

1. Some of our intentional mental states—for example, our desires, emotions, intentions, and beliefs—directly reflect certain judgments, evaluations, or appraisals. For example: fear involves the judgment that something is dangerous or threatening, resentment involves the judgment that I have been wronged, the desire to do X involves the judgment that X is good in some way (e.g., pleasant, fun, exciting), and so on. There is not just a causal connection between these evaluations and the attitudes in question; there is a conceptual connection between them. These attitudes are partially constituted by these evaluative judgments.
2. Some other intentional mental states—for example, what we notice and neglect, what does and does not occur to us, and what we see as relevant in our practical deliberations—indirectly reflect certain judgments or evaluative commitments. There is, if not a conceptual connection, at least a rational connection between these unreflective patterns of awareness and what we care about or regard as important or significant.
3. In virtue of these normative connections, these attitudes and patterns of awareness are the kinds of states for which reasons or justifications can appropriately be requested. That is to say, since these attitudes are linked with our judgment in various ways, we can, in principle, be asked to defend them and to give up or modify them if an adequate defense cannot be provided. (In the case of noticing, we can be criticized and asked to acknowledge fault for failing to notice something if this failure can reasonably be taken to reflect an [objectionable] judgment that the thing in question is not important.)
4. But “being responsible” for something is primarily a matter of being the appropriate target of these sorts of normative demands (and we can be appraised in various ways for how well or poorly we meet

these demands). Although the standards or expectations on which these demands are based will vary depending upon what is in question, the basic condition of moral attributability is the same.

5. Therefore, since we are open, in principle, to these sorts of demands with respect to our attitudes, they are attributable to us in the way required for ascriptions of responsibility in the most basic sense.

I have focused in this article on establishing the conditions of moral attributability and have said very little about the particular kind of appraisal or criticism that one must be open to in order to qualify as responsible for something in the relevant sense. One question that might be asked here is whether we are open to the very same kinds of appraisals for our attitudes as we are for our voluntary actions. An adequate defense of the rational relations view would need to address this question and would also need to say more about how this view would deal with the problems posed by recalcitrant attitudes, *akrasia*, and certain other forms of irrationality. For now, though, I hope at least to have made plausible the claim that what makes us responsible for our attitudes is not that we have voluntarily chosen them, endorsed them, or have voluntary control over them but that they are the kinds of states that reflect and are in principle sensitive to our rational judgments.