Philos Stud (2008) 138:367–392
DOI 10.1007/s11098-006-9048-x

ORIGINAL PAPER

Control, responsibility, and moral assessment

Angela M. Smith

Received: 15 June 2006 / Accepted: 23 October 2006 / Published online: 15 December 2006
© Springer Science+Business Media B.V. 2006

Abstract Recently, a number of philosophers have begun to question the commonly held view that choice or voluntary control is a precondition of moral responsibility. According to these philosophers, what really matters in determining a person's responsibility for some thing is whether that thing can be seen as indicative or expressive of her judgments, values, or normative commitments. Such accounts might therefore be understood as updated versions of what Susan Wolf has called "real self views," insofar as they attempt to ground an agent's responsibility for her actions and attitudes in the fact (when it is a fact) that they express who she is as a moral agent. As such, they seem to be open to some of the same objections Wolf originally raised to such accounts, and in particular to the objection that they cannot license the sorts of robust moral assessments involved in our current practices of moral responsibility. My aim in this paper is to try to respond to this challenge, by clarifying the kind of robust moral assessments I take to be licensed by (at least some) non-volitional accounts of responsibility and by explaining why these assessments do not in general require the agent to have voluntary control over everything for which she is held responsible. I also argue that the limited applicability of the distinction between "bad agents" and "blameworthy agents" on these accounts is in fact a mark in their favor.

Keywords Responsibility · Blame · Moral Assessment · Real Self Views · Accountability · Control

1 Introduction

Recently, a number of philosophers have begun to question the commonly held view that choice or voluntary control is a precondition of moral responsibility. According to these philosophers, what really matters in determining a person's responsibility

---

A. M. Smith (✉)
Department of Philosophy, University of Washington, Box 353350, Seattle, WA 98195, USA
e-mail: angsmith@u.washington.edu

for some thing is whether that thing can be seen as indicative or expressive of her judgments, values, or normative commitments.[1] In some cases (e.g., in determining a person's responsibility for a bodily movement), it may make sense to ask whether the agent has *voluntarily chosen* the thing in question, because that will determine whether that thing can reasonably be taken to express her judgments. But in other cases (e.g., in determining a person's responsibility for certain omissions, or for her desires, emotions, and beliefs), there does not seem to be any need to appeal to the agent's choices, because the thing in question can be seen as directly expressive of her judgments and normative commitments. According to these philosophers, the condition of voluntary control should be replaced by a condition of judgment-sensitivity (See Scanlon, 1998, chap. 6), or character-expressiveness (See Sher, 2006), or judgment-dependence (See Smith, 2005).

These accounts might be understood as updated versions of what Susan Wolf has called ''real self views,'' insofar as they attempt to ground an agent's responsibility for her actions and attitudes in the fact (when it is a fact) that they express *who she is* as a moral agent.[2] As such, they seem to be open to some of the same objections Wolf originally raised to such accounts, and in particular to the objection that they cannot license the sorts of robust moral assessments involved in our current practices of moral responsibility. The idea here is that while we may be open to various forms of moral and non-moral evaluation on the basis of what we are like as moral agents, we are not open to the particular kind of assessment that marks out the notion of moral responsibility as we ordinarily understand it—namely, moral praise and blame.[3] The mere fact that an action or attitude expresses something about ourselves is not enough to show that we are *responsible* for it, on Wolf's view; we must also show that it is our fault (or to our credit) that we are the way we are, that is, that we have *control* over who we ultimately are.[4] This, in turn, has led some philosophers to think that the kind of ''responsibility,'' and the corresponding kind of ''blame,'' we get from these updated real self views is ethically superficial: it is a kind of responsibility and blame that we can equally attribute to other living and non-living things on the basis of their good or bad qualities.

My aim in this paper is to try to respond to this challenge, by clarifying the kind of robust moral assessments I take to be licensed by (at least some) non-volitional accounts of responsibility and by explaining why these assessments do not in general require the agent to have voluntary control over everything for which she is held responsible. My strategy will be as follows: In the next section, I will provide a brief sketch of my own non-volitionalist account of responsibility, which I have elsewhere called the rational relations view. In section three I will examine Wolf's original critique of real self views and will look at a recent response to that critique by Gary Watson. Though Watson vigorously defends real self views against Wolf's charge of "superficiality," he nevertheless concedes that the kind of blame licensed by these accounts (which he calls "aretaic blame") is different from (and less serious than) the kind of blame ("accountability blame") that he says is the focus of most traditional accounts of moral responsibility. If this is correct, then it would turn out that the sort of non-volitional account of responsibility that I and others have sought to defend does not really represent a genuine alternative to traditional volition-based accounts of responsibility after all. For real self views would not be able to justify those central forms of assessment that are constitutive of our current moral practices. In section four, I will respond to this challenge, by explaining what I take to be distinctive about moral assessment as distinguished from certain other forms of evaluation. I locate the special force of moral appraisal in the idea that these assessments involve implicit demands for justification, demands that are made on the basis of moral standards we expect reasonable persons to accept. Moral blame and criticism are distinguished from other forms of assessment, on this view, both by the kinds of responses they call for (explanation, justification, or admission of fault) and by the particular standards to which these demands for justification appeal. Once we see what is distinctive about moral criticism, however, we will see that choice or voluntary control is not, in fact, a precondition of these forms of assessment. Finally, in section five, I will consider whether it is a strike against non-volitional accounts in general, and the rational relations view in particular, that they do not leave us much room to draw a distinction between "bad agents" and " blameworthy agents."[5] I will argue that the limited applicability of this distinction when it comes to the assessment of persons actually counts in favor of these accounts, as it amounts to a refusal to regard persons as the passive victims of their character and conduct.

---

[1] Robert Adams (1985) was one of the first to seriously question volitional accounts of responsibility. For other recent defenses of non-volitionalism, see Scanlon (1988, 1998); Sher (2005, 2006); and Smith (2000, 2005).

[2] Wolf discusses what she calls ''Real Self Views'' in Chapter 2 of Wolf (1990). Gary Watson and R. Jay Wallace have both discussed Wolf's view at length in their own work. See Watson (1996), and Wallace (1996), pp. 52–62.

[3] For a recent critique of my own account on these grounds, see Neil Levy (2005). For similar claims, see Blum (1980), p. 189; Slote (1992), pp. 120–124; and Wallace (1996), pp. 126–127. Justin Oakley draws a distinction between assessments that presuppose responsibility (blameworthiness and praiseworthiness) and assessments that do not (estimability and disestimability), and argues that we can be subject to both forms of appraisal on the basis of our attitudes (in particular, for our emotions). In order to be subject to the former assessments, however, we must have made or culpably failed to make voluntary efforts to cultivate our emotional dispositions. See Oakley (1992), chs. 4 and 5. For a similar view, see Sankowski (1977), p. 838.

[4] Wolf (1990), p. 44.

2 The rational relations view

Before turning to Wolf's challenge to real self accounts of responsibility, it might be helpful to give a brief summary of the particular sort of non-volitional view I seek to defend. This view, which I have elsewhere called the rational relations view, makes rational judgment rather than choice or voluntary control the basic condition of moral responsibility.[6] To say that an agent is morally responsible for some thing, on this view, is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it.[7] Volitionalists about moral responsibility, by contrast, believe that a person is morally responsible only for those things that she has—directly or indirectly—chosen.[8] The rational relations view is thus committed to extending the scope of our moral responsibility to include many things that are commonly excluded by volitionalist accounts. For example, most of our desires, emotions, beliefs, and other attitudes seem to meet this condition of judgment-dependence, even though they do not commonly reflect a choice or decision, and are not normally under our voluntary control. These states are ‘‘judgment-dependent’’ in the sense that they generally reflect and are sensitive to our (sometimes hasty, mistaken, or incomplete) judgments about what reasons we have, and they are generally responsive to changes in these judgments. We are ‘‘responsible for’’ these things, therefore, because they reflect rational assessments for which we are appropriately regarded as answerable. Our patterns of awareness—e.g., what we notice and neglect, and what occurs to us—can also be said to reflect our judgments about what things are important or significant, so these responses, too, will count as things for which we are responsible on the rational relations view. What matters, on this account, is whether an action or attitude is normatively connected to a person's underlying judgments in such a way that she can, in principle, be called upon to defend it with reasons and to acknowledge fault if an adequate defense cannot be provided. Bodily movements and mental states that are not even in principle answerable to a person's judgment are therefore not the sorts of things for which we are responsible, on this account; but we are responsible for most of our desires, emotions, beliefs, and other attitudes, despite the fact that they do not generally arise from conscious choice or decision and are not normally under our immediate voluntary control.

---

[5] This is the central objection Levy (2005) raises against non-volitional accounts.

[6] Smith (2005). For related discussion and defense, see Smith (2000) and (2004).

[7] I say appropriate "in principle" because whether it would be appropriate "in fact" for any particular individual to make such a demand will often depend upon contextual factors that have nothing to do with a person's responsibility for the action or attitude in question. I explore these issues further in Smith (forthcoming, 2007).

[8] Levy (2005) provides an explicit statement and defense of this volitional view.

It is important to emphasize that the rational relations view and the volitional view are intended to be competing accounts of the same thing: namely, the conditions that must be met in order for someone to count as morally responsible for an action or an attitude. And on both accounts, to say that someone is ‘‘morally responsible’’ for some thing is to say that she is open, in principle, to moral appraisal—including moral praise and blame—on the basis of it (where nothing is implied about what that appraisal, if any, should be).[9] Where these accounts differ is in their understanding of the kind of connection a person must have to an action or attitude in order for her to be morally responsible for it in the relevant sense.

---

[9] In his critique of non-volitionalist accounts, Levy (2005) provides a similar, but slightly different, account of what it means to say that an agent is morally responsible: ‘‘To say that an agent is morally responsible (for an act, omission or attitude) is to say that the Strawsonian reactive attitudes are justified in relation to her with regard to that act, omission or attitude,’’ p. 2. I have two minor concerns about this formulation: first, this formulation suggests that an agent is morally responsible only for those actions or attitudes that in fact warrant a reactive attitude (hence only for those actions/attitudes that justify attitudes of esteem or indignation or some other reactive attitude). But this would mean we are not morally responsible for most of our boring, neutral, everyday actions, actions that do not warrant any particular reactive attitudes on the part of ourselves or others. We don't want an account of moral responsibility that implies that we are only morally responsible for those actions/attitudes that in fact warrant either praise or blame. Thus openness ‘‘in principle’’ to moral appraisal (or to the Strawsonian reactive attitudes) is what matters. Second, the question of whether a Strawsonian reactive attitude is ‘‘justified’’ in a particular context might involve considerations that have nothing to do with the agent's responsibility and blameworthiness (e.g., we might think that someone who shares a fault with the agent might not be ‘‘justified’’ in feeling indignation toward her, yet that seems irrelevant to the question of whether she is in fact responsible and blameworthy for the thing in question). For this reason, I think it is less problematic to formulate the notion of what it is to be morally responsible for some thing in terms of whether the person is open, in principle, to moral appraisal on account of the thing in question, rather than in terms of whether a Strawsonian reactive attitude is justified. I don't think this affects anything of substance in the debate between volitionalism and non-volitionalism.

Volitionalists claim that a person must have *voluntarily chosen to bring about* an action or attitude in order for her to be responsible for it; the rational relations view claims that a person's judgment must be *reflected in* an action or attitude in order for her to be responsible for it.

The worry that has been raised against non-volitionalist accounts of this sort is that they cannot, in fact, license the particular sorts of appraisals constitutive of our practices of moral responsibility. In particular, it has been argued that we cannot regard a person as *blameworthy* simply for having an attitude that she has not chosen and that is not under her voluntary control. Perhaps we can say that a person who has certain attitudes is morally "flawed" or "defective" or "bad," but it only makes sense to say that she is *blameworthy* for these flaws if she has voluntarily brought about these conditions in herself. This is the challenge to which I now turn.

3 Responsibility as self-disclosure

As I have just suggested, most non-volitionalist accounts of moral responsibility appeal to the fact that, in normal circumstances, our actions and attitudes can be seen as dependent upon and expressive of our underlying judgments and evaluative assessments. This focus on the expressive or self-revealing character of our actions and attitudes goes naturally with a conception of responsibility that Gary Watson has called "the self-disclosure view" (Watson, 1996, p. 228). Watson describes this conception of responsibility in the following way:

> The self-disclosure view describes a core notion of responsibility that is central to ethical life and to ethical appraisal. In virtue of the capacities identified by the self-disclosure view, conduct can be attributable or imputable to an individual as its agent and is open to appraisal that is therefore appraisal of an individual as an adopter of ends. Attributability in this sense *is* a kind of responsibility. In virtue of the capacities in question, the individual is an agent in the strong sense, an author of her conduct, and is in an important sense answerable for what she does (Watson, 1996, p. 229).

Watson's focus here is on responsibility for actions, and therefore the "capacity" he highlights is our capacity to set ourselves ends. But it should be clear, I hope, how such a view might be extended to give a notion of responsibility for attitudes.[10] Insofar as we have the capacity to recognize, assess and respond to reasons counting in favor of certain attitudes as well as reasons counting in favor of certain actions, our attitudes can be seen as "imputable to us as agents" in much the same way that our actions can be.

One of Watson's aims in this article is to defend this conception of responsibility as self-disclosure against a challenge put forward by Susan Wolf in her book *Freedom Within Reason*. Wolf argues there that self-disclosure views, or what she calls "real self views," can account at best for a "superficial" kind of responsibility, a kind of responsibility that is merely causal and can be applied to "[e]arthquakes, defective tires, and broken machines...[as well as] dogs and children and adults with various physical and mental handicaps'' (Wolf, 1990, p. 40). All of these objects and agents, she suggests, can have ''good or bad qualities'' which can be expressed in ''welcome or unwelcome'' ways. But to say that is not yet to say that they are responsible for these qualities in the ''deep'' way that we ordinarily consider normal adult human beings to be responsible for their own character or conduct. As she puts the point,

---

[10] Watson himself does not address this question, and I do not know whether he would endorse this extension.

> When we say that an individual is responsible for an event in the superficial sense, we identify the individual as playing a causal role that, relative to the interests and expectations provided by the context, is of special importance to the explanation of that event. And when we praise or blame an individual in the superficial sense, we acknowledge that the individual has good or bad qualities, or has performed good or bad acts. But when we hold an individual morally responsible for some event, we are doing more than identifying her particularly crucial role in the causal series that brings about the event in question. We are regarding her as a fit subject for credit or discredit on the basis of the role she plays. When, in this context, we consider an individual worthy of blame or praise, we are not merely judging the moral quality of the event with which the individual is so intimately associated; we are judging the moral quality of the individual herself in some more focused, noninstrumental, and seemingly more serious way. We may refer to the latter sense of responsibility as deep responsibility, and we may speak in connection with this of deep praise and blame (Wolf, 1990, pp. 40–41).

As Watson points out, Wolf's objection in this passage is two-fold: first, that responsibility on the self-disclosure view is ''merely causal (and therefore ethically superficial)''; and second, that the kinds of appraisals to which we are open in virtue of this kind of responsibility ''are simply descriptions of a thing's qualities and differ in kind from moral blame in the strict (deep?) sense'' (Watson, 1996, p. 232). It is this second criticism that I want to focus on in the remainder of this section, since it poses an important challenge to views like mine according to which an agent is responsible for something just in case that thing can be seen as expressive of her evaluative judgments—expressive of her ''real self,'' as Wolf would put it. If this condition of moral attributability licenses only this superficial, merely descriptive, form of ethical appraisal, then it would be clear that I have not really succeeded in offering an alternative account of the conditions of moral responsibility as this concept is ordinarily understood.

3.1 Deep and superficial assessment

In order to assess Wolf's objection, we first need to gain a clearer understanding of the distinction she invokes between ''deep'' versus ''superficial'' forms of assessment. Of course, part of the dispute between Wolf and Watson is over which forms of assessment can be characterized as ''deep,'' so we do not want to prejudge that issue at the outset. But I think there are clear enough examples of each kind of assessment at the extremes to give us a sense of the general character of the distinction, even if we are unsure about where, precisely, to draw the line between them.

We might start, then, by considering some cases of clearly "superficial" forms of assessment. The average elementary school classroom is filled with examples of this sort. When I was growing up, I was always one of the shortest kids in my class. Each year the school's doctor would come in and measure all of us, and I would always end up in the "below average" portion of the age-height chart. This was, of course, deeply mortifying for me at the time. But the assessment "below average," as applied to a physical characteristic such as one's height, is "superficial" in an obvious sense. It does not carry with it any implication of fault, nor does it provide any grounds for disapproval or reproach; it is merely a description of where one falls relative to one's peers in one's physical development.

In addition to measuring our height, the doctor would also administer sight and hearing tests. Although I always passed the eye exams with flying colors, I once failed the hearing test. This was even more deeply mortifying than ending up below average on the height chart, since it involved actually "failing" a test that most of my peers seemed to have no difficulty passing. But here again, the "failure" in question was of a superficial sort. It did not indicate any fault on my part, or provide any grounds for disapproval or reproach; it merely indicated that I fell below a certain standard in one of my basic physical abilities. In Wolf's terms, we can acknowledge that having poor hearing is a "bad" thing, without claiming that a person who has this trait is a "fit subject for discredit."

Assessments of people as beautiful or ugly, intelligent or stupid, and athletic or unathletic, likewise, seem to be "superficial" forms of evaluation. Though others may (cruelly) mock or despise us for our basic physical or intellectual attributes, it seems they cannot properly fault or reproach us for them. It would be inappropriate, under any circumstances, for someone to "blame" me for having big ears, for example, or to be indignant about my low IQ. These things simply cannot serve as legitimate grounds for moral or non-moral discredit.

In his well-known article, "Free-will, Praise, and Blame," J. J. C. Smart refers to evaluations of this sort, which merely report how a thing or person stands with regard to some standard, as "grading" evaluations. The opposite of "praise" when used in this grading sense is not blame, but what Smart calls "dispraise." As he puts it,

> When we praise Smith for his mathematical talent we do not imply that we blame Jones because, try as he may, he cannot handle *x*'s and *y*'s. Of course we may well say that a girl is ugly, a footballer incompetent, or a man unmathematical, and this is the opposite of praise. But it is not blame. Praise and dispraise, in this sense, is simply grading a person as good or bad in some way (Smart, 1961, p. 303).

Smart goes on to suggest that, "In this sense of 'praise' we may praise moral qualities and moral actions in exactly the same way as we may praise beauty, intelligence, agility, or strength" (Smart, 1961, p. 304), and we may 'dispraise' moral qualities and actions in the same way. 'Blame', in his view, should be understood as 'dispraise' of exactly this sort, with the added "proviso" that it is directed at a feature that can be influenced by unwelcome evaluations of this sort. We "blame" a lazy child for not doing his homework, he says, but not an exceedingly stupid one, simply because "the lazy boy can be influenced in such ways" while the stupid boy cannot (Smart, 1961,

p. 302). But 'blame' in this sense does not bring with it any "deeper" sense of responsibility, or any further implication of fault or discredit.[11]

Smart, then, embraces precisely the "superficial" sense of responsibility and assessment that Wolf argues does not adequately capture our ordinary practices of responsibility ascription. He freely acknowledges that most people do not in fact praise and blame people "in this dispassionate and clear-headed way" (Smart, 1961, p. 305); but he insists that they should. If we were clear about what we are doing, he claims, we would see that the only difference between "moral blame" and unwelcome grading evaluations is that the former can be instrumentally useful in bringing about traits or dispositions in others that are socially beneficial. His theory of responsibility is, for this reason, deeply revisionist.

Most people, however, agree with Wolf that there is an important difference between grading a person's hearing or physical attractiveness, on the one hand, and judging her morally, on the other, a difference that cannot be accounted for in the purely instrumental or pragmatic way that Smart proposes. This, presumably, is what lies behind Wolf's claim that "when we consider an individual worthy of blame or praise, we are not merely judging the moral quality of the event with which the individual is so intimately associated; we are judging the moral quality of the individual herself in some more focused, noninstrumental, and seemingly more serious way" (Wolf, 1990, p. 41).

When we blame someone for a cruel action or attitude, for example, we do not seem to be saying merely that she has a quality that fails to meet a certain objective standard of moral goodness (as my hearing failed to meet an objective standard of aural goodness); we seem to be saying that *she* has failed in some way, and that she is open to serious moral criticism for this failure. We take her to be *answerable* for such a failure in a way that we do not normally consider people to be answerable for their aural abilities or physical attractiveness. Saying that someone is "answerable" here, moreover, is not simply a way of saying that we regard the trait or quality in question as "modifiable" in the way Smart's view suggests.

I will say more about how I think we should understand the particular kind of "depth" associated with moral (and some kinds of non-moral) appraisal in section four. At this point, however, I am simply trying to identify some of the central differences between "deep" and "superficial" forms of assessment. The comparison with Smart has, I think, brought out three of these differences. First, both Wolf and Smart seem to agree that there is a difference between ‘‘grading’’ someone on the basis of a (moral) quality and ‘‘judging’’ him on the basis of it. The former superficial assessment is merely descriptive, while the latter deep assessment goes beyond mere description in some important sense. Second, they seem to agree that deep assessment involves ascriptions of credit or fault, while superficial assessment does not. Finally, they seem to agree that deep assessment normally supports certain attitudes and emotional responses (e.g., indignation, resentment, and esteem), while superficial assessment does not.[12] Where they differ is that Wolf thinks we can give an account of the conditions of responsibility that justifies these features of ‘‘deep’’ assessment, while Smart thinks we cannot.

---

[11] This understanding of the nature and purpose of moral appraisal did not originate with Smart, of course. David Hume's denial that there is any "deep" difference between the appraisal of natural abilities and the appraisal of moral qualities in Book III, Part 3, Section iv of *A Treatise of Human Nature* involves a similar understanding of moral praise and blame: "Men have observ'd, that tho' natural abilities and moral qualities be in the main on the same footing, there is, however, this difference betwixt them, that the former are almost invariable by any art or industry; while the latter, or at least, the actions, that proceed from them, may be chang'd by the motives of reward and punishment, praise and blame. Hence legislators, and divines, and moralists, have principally applied themselves to the regulating these voluntary actions, and have endeavour'd to produce additional motives for being virtuous in that particular. They knew, that to punish a man for folly, or exhort him to be prudent and sagacious, wou'd have but little effect; tho' the same punishments and exhortations, with regard to justice and injustice, might have a considerable influence." See Hume (1739, 1978), p. 609. Hume places less emphasis on this utilitarian aspect of his theory in Appendix IV of his *Enquiry Concerning the Principles of Morals* although he still insists that there is not a deep distinction to be drawn between virtues and talents, vices and defects. See Hume (1751, 1975), pp. 312–323. For other defenses of this view, see Schlick (1939); and, more recently, Daniel Dennett (1984), ch. 7.

Wolf’s challenge to self-disclosure or attributability views, then, is that they can at best license the kind of superficial appraisal Smart articulates, but not deep moral praise and blame as these notions are ordinarily understood. This is no problem for Smart, of course, since he thinks the ‘‘ordinary’’ notion of praise and blame is hopelessly confused anyway. But this is a serious challenge to a theory such as mine, which purports to be offering an account of responsibility that is internal to our moral concepts as we now understand them. Before giving my own reply to Wolf’s challenge, however, I want to look briefly at Watson’s response to it. Watson’s analysis is helpful, because it brings out the complexity of the issues surrounding questions of moral responsibility in a particularly clear way. Although I believe his attempt to draw a distinction between what he calls ‘‘aretaic appraisal’’ and ‘‘accountability appraisal’’ is ultimately untenable, it highlights a number of important ambiguities in our thinking about moral appraisal that it will be helpful to sort through before presenting my own account.

3.2 Aretaic appraisal and accountability appraisal

Watson’s basic thesis in ‘‘Two Faces of Responsibility’’ is that there are two different perspectives we can take on questions of responsibility, one of which is ‘‘prompted by a concern with agency or attributability,’’ and the other by a concern with ‘‘accountability’’ (Watson, 1996, p. 229). The former perspective he describes as the ‘‘aretaic perspective,’’ and he suggests that the nature and conditions of ‘‘aretaic evaluation’’ may be different from and less stringent than the conditions that must be met in order for a person to be regarded as ‘‘accountable’’ for her attitudes or conduct. Thus he concedes to Wolf that self-disclosure views might not be able to account for all of the assessments involved in our ordinary moral practices. Nevertheless, he insists that aretaic evaluations are ‘‘deep’’ evaluations, in Wolf’s sense, because they ‘‘implicate one’s practical identity’’ (Watson, 1996, p. 234), and are thus ‘‘inescapably evaluations of the agent’’ (Watson, 1996, p. 233).

Watson tries to bring out the difference between these two perspectives with the use of an example, drawn from Peter van Inwagen’s *An Essay on Free Will* (1980, p. 207). Van Inwagen writes of a colleague who claims to deny the reality of moral responsibility, but who, upon learning that some of his books were stolen, says "That was a shoddy thing to do." Van Inwagen claims that in asserting this, his colleague betrays the fact that he really does believe in moral responsibility after all, because "no one can consistently say that a certain act was a shoddy thing to do and say that its agent was not morally responsible when he performed it" (van Inwagen, 1980, p. 207).

---

[12] Smart (1961) thinks that blame should be ‘‘just as dispassionate as dispraise of a woman’s nose,’’ p. 305. Indeed, a large part of the motivation for Smart’s paper, it seems, is to combat what he sees as the pernicious assumption that ‘‘righteous indignation is an appropriate emotion in certain circumstances’’ (p. 291). There is some irony, therefore, in his own righteous indignation toward those ‘‘pharisees’’ who continue to ‘‘judge’’ others in spite of the alleged metaphysical facts of the matter.

In response, Watson proposes that the kind of responsibility the colleague claims to reject, and the kind of responsibility his comment seems to commit him to, may in fact be different. And if this is so, he may not be guilty of a contradiction in the case. He puts the point in the following way:

> To say (or judge) that stealing the books was a shoddy thing to do implies that the thief behaved shoddily, and hence that the act was attributable to the agent. To call the conduct "shoddy" is to see it as "inferior goods," as a poor exercise of human evaluative capacities, as characteristic of someone who cares little about standards of excellence in human affairs....However, such attributions should not be too quickly grouped with other judgments which the colleague allegedly wishes to eschew (Watson, 1996, p. 230).

The "other judgments" which van Inwagen's colleague may mean to be rejecting, Watson suggests, are those associated with "the practice of moral accountability":

> But 'shoddy' need not express "censure." That implies a public forum, in which the subject is liable to formal sanction. To speak of conduct as deserving of "censure," or "remonstration," as "outrageous," "unconscionable" (and on some views, even as "wrong"), is to suggest that some *further response* to the agent is (in principle) appropriate. It is to invoke the practice of holding people morally accountable, in which (typically) the judge (or if not the judge, other members of the moral community) is entitled (in principle) to react in various ways (Watson, 1996, p. 230).

Watson concludes from all this that blaming judgments of the sort made by van Inwagen's colleague "do commit him to attributability of the kind identified by the self-disclosure view. But nothing in this view explains or justifies any such reactive entitlement" (Watson, 1996, p. 230). He goes on to say that aretaic evaluations "are independent of the practices of moral accountability. They invoke only the attributability conditions, on which certain appraisals of an agent are grounded" (Watson, 1996, p. 231), and that these aretaic evaluations must be distinguished from "accountability blame," which "involve[s] the idea that agents deserve adverse treatment or 'negative attitudes' in response to their faulty conduct" (Watson, 1996, p. 230).

Watson's suggestion here is that there is a distinctive form of moral appraisal, which he calls "aretaic blame," that falls in between mere grading evaluations (Smart's "dispraise") and full-blooded moral blame ("accountability"). Aretaic blame, he claims, is a deep form of appraisal, because it involves the evaluation of an agent as the "author" of her conduct. But evaluating someone as a bad "author," he suggests, is different from holding her "accountable" for the story she tells. If we hold her "accountable" for her story, we are saying that she is a legitimate target of certain forms of adverse treatment on the basis of the story she tells. But in order for someone to be the legitimate target of adverse treatment, it is not sufficient that her attitudes and actions reflect her judgment (that she be "the author" of her conduct);

she must also have had a fair opportunity to avoid being subject to that adverse treatment (Watson, 1996, p. 237). Therefore, he concludes, the conditions of "accountability blame" go beyond the conditions of "aretaic blame," and self-disclosure views simply do not justify many of the activities and responses associated with our current practices of moral responsibility.

Watson's discussion is very helpful for bringing out some of the different concerns we might have in mind when we approach questions of moral responsibility. I have a suspicion, however, that the distinction Watson draws between "aretaic blame" and "accountability blame" rests on an ambiguity in the notion of "accountability" which, once cleared up, will show not that there are two different senses of *responsibility* or *moral blameworthiness* at work, but rather that there is a distinction between judging a person to be responsible and culpable for her behavior, on the one hand, and engaging in various forms of blaming activity (including punishment), on the other. But the fact that there may be further conditions governing our blaming activities does not, I believe, support the claim that attributability views are working with a different and weaker conception of the relevant forms of moral appraisal. Since Watson's discussion gives this impression, I think it would be worthwhile to examine his proposal that there is a distinctive form of moral appraisal, "aretaic blame," which falls in between mere "dispraise" and full-blooded moral blame.

We should start, then, by considering Watson's reflections on the use of the term 'shoddy'. He makes three points about how this aretaic form of blame differs from accountability blame: first, 'shoddy' need not express "censure"; second, censure "implies a public forum in which the agent is liable to formal sanction"; and third, "to speak of conduct as deserving of 'censure,' or 'remonstration,' as 'outrageous,' 'unconscionable' (and on some views, even as 'wrong'), is to suggest that some *further response* to the agent is (in principle) appropriate." Consider the claim that 'shoddy' need not express censure. If by "censure" is meant a judgment involving condemnation, it seems to me it is hard to deny that van Inwagen's colleague is expressing an attitude of censure toward the thief's conduct. 'Shoddy' in this context implies a certain meanness or contemptibility, a despicable lack of concern for the interests of others; one cannot apply such an epithet to a person's conduct, it seems, without expressing some degree of condemnation for it.

Watson might dispute this. In his gloss on the colleague's comment, he uses expressions such as "inferior goods," "a poor exercise of human evaluative capacities," and lack of care for "standards of excellence in human affairs." Perhaps one can evaluate something as "shoddy," as failing to meet certain standards of excellence, without expressing anything properly called "condemnation" for it at all. To call a certain piece of human behavior "shoddy," then, would simply be to grade it as "inferior goods" in exactly the same way that to call a piece of manufacturing work "shoddy" is to grade it as "inferior goods." Such an assessment need not involve condemnation, and therefore need not involve the idea that the person in question is at fault in some way for his or her conduct.

On this interpretation, however, the notion of aretaic appraisal would seem to be very close to Smart's idea of praise and dispraise. Aretaic appraisals would report how a person stands with regard to certain standards of human excellence, but would not bring with them any further implications of fault or discredit, and would not support any demands for reasonable regard on the part of others. But if aretaic evaluations are simply "grading" evaluations, which do not imply any normative failure on the part of the person appraised, then Watson's insistence that they are nevertheless "deep" forms of assessment becomes considerably more dubious. It does not seem to help at this point to insist that, when applied to people, these assessments concern a person's "practical identity," for all that would seem to follow from this is that some practical identities can be graded more highly than others. If this is the only kind of appraisal we can make of persons on self-disclosure or attributability views, then I think Wolf is right to complain that such views can account at best for a "superficial" kind of responsibility.

But perhaps Watson would be willing to concede that the use of the term 'shoddy' in this context does involve "censure" in the sense of a *judgment* of condemnation. His claim that censure "implies a public forum in which the subject is liable to formal sanction" suggests that when he talks of "censure" he has in mind something stronger than a judgment of condemnation, something involving an official reprimand or punishment of some sort. If so, then he might be correct in saying that van Inwagen's colleague need not be committed to the claim that it would be a good thing for the person who stole his books to be publicly and formally sanctioned for his shoddy behavior. But now we need to ask what the relation is between judgments of moral responsibility and official public practices of censure and reward. If I behave in a "shoddy" way toward a friend, say by spreading malicious rumors about her behind her back, it seems to me that I am both responsible and open to serious moral criticism for my conduct. And yet, I do not think I am liable to "formal sanction," if we mean by this official public censure or punishment. But that seems to have no bearing whatsoever on the question whether I am responsible and open to serious moral criticism for my conduct.

At this point Watson might object that I am taking his words too literally. What matters is not that there be some "official" authority, or that there be an actual public setting for the envisioned drubbing to take place. What distinguishes aretaic appraisal from accountability appraisal is simply the fact that the latter form of assessment involves the idea that "some *further response* to the agent is (in principle) appropriate." In order to assess Watson's claim, however, we need to know what these "further responses" are. If the "further responses" include official public censure and other forms of punishment, then he would be right to claim that the conditions for aretaic blame and the conditions for accountability blame are different. But that would be because "accountability blame," interpreted in *this* way, involves forms of ill-treatment that carry a heavier burden of justification than the kind of blame that is in question when we are discussing the basic conditions of moral responsibility. ("Accountability blame," understood in this way, is akin to legal punishment).

The real difficulty with Watson's account, however, is that he does not restrict the relevant "further responses" to official public censure and punishment, but includes "negative attitudes" and even various judgments, such as that a person's conduct is "outrageous," "unconscionable" or even "wrong." But if the mere judgment that a person's conduct is wrong or objectionable counts as one of the relevant "further responses," then this surely cannot be what distinguishes "accountability blame" from "aretaic blame." Or rather, if it *is* what distinguishes them, then again it seems aretaic appraisals must simply be forms of praise and dispraise. If aretaic appraisals cannot even be said to involve judgments of *this* sort, if "shoddy" cannot even be taken to imply "wrong" (or at least "objectionable"), then again it is unclear in what sense aretaic blame can be considered ‘‘deep’’ (or even a form of ‘‘blame,’’ for that matter).

In the end, I think Watson is conceiving of ‘‘accountability blame’’ as active blaming responses of various sorts, including both public censure and punishment as well as the active expression of negative attitudes such as blame, criticism, resentment, and indignation. But it is important to recognize that ‘‘accountability blame,’’ interpreted in this way, is a type of activity that is governed by a variety of moral and non-moral norms, a great many of which have nothing to do with the basic conditions of moral responsibility. Our judgments about when it would or would not be appropriate to actively express moral criticism or to punish someone for some action or attitude are sensitive to a host of considerations that have little or nothing to do with that person’s responsibility and culpability for the action or attitude in question.[13] These considerations can include, among other things, our relation to the person, our stake in the matter, the significance of fault, and the person’s own response to her failure. We may conclude on the basis of any of these things that blaming activities and responses would be inappropriate in a given case, without thinking that the person is therefore not responsible (or not ‘‘open’’ to moral criticism) for the action or attitude in question.

I have gone on about this at some length, because I believe these ambiguities have generated a great deal of confusion in the literature on moral responsibility, and in particular in discussions of responsibility for attitudes. From the fact that it would often be inappropriate for us to express moral criticism or to sanction people for their desires, emotions, and other attitudes, it is sometimes inferred that they cannot *be* morally responsible for them. But this inference is too quick. It may be that the inappropriateness of sanctioning people in these cases has nothing to do with issues of responsibility or culpability, but with more general moral issues concerning how it is appropriate to treat people in light of their moral faults. We face exactly the same sorts of questions when it comes to determining how to treat people in light of their wrongful actions. These are important substantive moral questions, but they go beyond the basic questions of responsibility and moral culpability that I take to be at issue in these debates over the conditions of moral responsibility.

In saying this, I do not at all mean to be suggesting that questions about the appropriateness of particular forms of blaming behavior are unimportant, or that Watson is confused in thinking that these questions properly belong in discussions of moral responsibility. Indeed, I think Watson has done an admirable job of calling our attention to these issues and of exploring their complexity. My worry is simply that by talking about this in terms of two different senses of ‘‘appraisal’’ and correspondingly two different senses of ‘‘responsibility,’’ his analysis gives the impression that self-disclosure or attributability views are not, in fact, concerned about giving the conditions of moral responsibility as this concept is ordinarily understood. But that is precisely what I and others who have recently defended non-volitional views take ourselves to be doing. In the next section, I will try to show that the rational relations view is working with full-blooded notions of moral appraisal, blame, and criticism. When it comes to *any* account of responsibility, however, there will always be further questions that can be asked about whether, when, and in what way negative moral appraisal should be expressed or acted upon.

---

[13] I defend this view more fully in Smith (forthcoming, 2007). For a similar view about the importance of distinguishing judgments of culpability and actual blaming responses, see the introduction to Fischer and Ravizza (1993), pp. 1–41. According to what they describe as a modified Strawsonian theory, ‘‘judging whether and to what degree to praise or blame someone involves a set of considerations over and above the base assessment that the person is rationally accessible to this sort of reaction,’’ p. 19.

4 Moral appraisal and the demand for justification

In the last section, I expressed some doubts about whether Watson's notion of ‘‘aretaic appraisal'' is ‘‘deep'' enough to put to rest Wolf's worry that self-disclosure or attributability views can at best account for a ‘‘superficial'' kind of responsibility. Some of the things Watson says about this form of appraisal make it sound like Smart's ‘‘praise and dispraise,'' which would lend support to Wolf's criticism, while other things he says make it sound more robust. Rather than trying to sort through these ambiguities in Watson's notion of aretaic appraisal, however, I want to turn now to the task of giving a fuller account of my own conception of moral appraisal. I believe this conception captures all of the ‘‘deep'' features Wolf associates with our ordinary concepts of moral praise and blame, and that the rational relations view provides a compelling account of the conditions which make these forms of assessment appropriate.

I want to begin, however, by looking again at Watson's very general suggestion that the kind of appraisal that is at issue in traditional debates about moral responsibility goes beyond the mere ascription of a negative or positive character trait to a person, and involves the idea that ‘‘some *further response* to the agent is (in principle) appropriate.'' He speaks of this as a ‘‘reactive entitlement,'' and claims that there is nothing in self-disclosure or attributability views that ‘‘explains or justifies any such reactive entitlement'' (Watson, 1996, p. 230). Now, I have already raised some questions and doubts about how Watson is conceiving of the ‘‘further responses'' associated with this reactive entitlement, so it is difficult to assess his claim about the alleged inability of attributability views to justify it. Setting these specific questions about Watson's view aside, however, I want to suggest that he *is* correct in claiming that some such reactive entitlement is implicit in our ordinary notion of moral appraisal, and that this is the key to understanding the particular kind of ‘‘depth'' which characterizes this form of assessment. The challenge, then, is to provide an accurate account of the nature and grounds of this reactive entitlement.

How, then, should we understand the ‘‘reactive entitlement'' implicit in moral appraisal? We should begin by looking more carefully at what is distinctive about moral appraisal—why, that is, moral praise and blame, unlike assessments of beauty or native intelligence, seem to go beyond mere unwelcome description, and involve something that might be called a reactive entitlement. I will focus, as most philosophers do, on the case of moral criticism, but I will try to note at various points how my understanding of moral appraisal can be extended to cover cases of positive moral assessment, as well.

The first thing to note is that moral criticism, unlike many other forms of negative assessment, seems to imply something about our *activity* as rational agents. To accuse someone of ‘‘selfishness,'' for example, is not simply to attribute a negative quality to her (like ugliness, or lack of intelligence), but to make a claim about her agential activity. It is to claim that she has failed, either in general or in a particular instance, to give proper weight or significance to the needs and interests of others in her attitudes and actions.[14] Moral criticism in general, I would argue, can only be directed to a person with regard to things that involve her rational activity in some way. Our physical traits and natural abilities, in themselves, do not seem to reflect our rational agency or activity in a way that would make them an appropriate basis for moral criticism.

This is related to the second important feature of moral appraisal: Moral criticism, by its very nature, seems to address a *demand* to its target. It calls upon the agent to explain or justify her rational activity in some area, and to acknowledge fault if such a justification cannot be provided.[15] If I am called ugly, I may feel embarrassed, amused, or incredulous, but what I am not likely to feel is challenged: that is, I will not see any reason to try to defend myself against the charge (or to apologize or make excuses for my alleged unsightliness). If I am called selfish, by contrast, I cannot regard that as merely an unwelcome description of some trait or feature I possess: it is a direct challenge to me as a moral agent, and therefore implicitly calls for some kind of response on my part. This is not to say, of course, that I will necessarily acknowledge or take seriously such a challenge; it is a point, rather, about the nature of moral appraisal itself, and how it differs from mere negative description.

I believe it is this feature of moral criticism—that it addresses a justificatory demand to its target—that constitutes the relevant ‘‘reactive entitlement’’ that Watson regards as essential to our practices of moral accountability. The ‘‘further response’’ to an agent that is (in principle) appropriate when we consider an agent morally responsible for some thing is that she can be asked to justify that thing, and to re-assess it if an adequate justification cannot be provided. If this is correct, then what primarily distinguishes moral criticism from other forms of negative description or assessment is that it involves this implicit demand for justification.

These features of moral appraisal—the fact that it implies something about a person’s activity as a moral agent, and the fact that it addresses a justificatory demand to its target – together imply that it is appropriately directed only at features of a person that can be said to reflect her practical agency. The key question at issue between volitionalist and non-volitionalist accounts of responsibility, then, is whether a person’s ‘‘practical agency’’ should be confined to her deliberate choices, or whether it extends more broadly to include her rational judgments and assessments quite generally. Many philosophers seem to think that it is only appropriate to direct moral *criticism* (as opposed to mere unwelcome evaluation) to features of a person that reflect a voluntary choice of some sort (See, for example, Levy (2005);

---

[14] In the case of positive moral appraisal, the claim is that the person has gone ‘‘above and beyond’’ what is morally required of her in some area. To praise someone as ‘‘generous’’ or ‘‘thoughtful,’’ for example, is to claim that she regularly or in some particular instance has given extraordinary attention to the needs and interests of others in her attitudes and actions.

[15] This view of what differentiates ‘‘deep’’ moral appraisal from mere unwelcome descriptions has been defended by Scanlon (1988), who claims that ‘‘Adverse moral judgment...differs from mere unwelcome description because it calls for particular kinds of response, such as justification, explanation, or admission of fault,’’ p. 171. See also Scanlon (1998), pp. 271–272. This feature of negative moral appraisal is more difficult to extend to its positive analogue, but I think we can say something like the following: positive moral appraisal, though it does not itself address a demand to its target, by its nature involves an acknowledgment that the agent has exceeded in some way the legitimate moral demands that apply to her. This is still very different from mere positive description, because it implies activity of some sort on the part of the agent in question.

Wallace (1996), pp. 128–133). Thus, unless we can show that a person's bad attitude reflects or is a result of her own choices, we cannot morally *criticize* her for this trait (though we can say that it is a ''bad'' trait for her to have). On the rational relations view, by contrast, moral criticism is appropriately directed to any features of a person that can be said to reflect her judgments or evaluative assessments, regardless of whether we think those features reflect or have resulted from a voluntary choice on her part.

How, then, are we to adjudicate this dispute between the volitionalist and the non-volitionalist? Can it be appropriate to address a ''demand'' to an agent to justify something she has not deliberately chosen, and over which she may not have direct voluntary control? Does this not violate the well established ''ought implies can'' principle?[16]

These are large and important questions, which deserve a much fuller treatment than I will be able to provide here.[17] But as a start, I think it useful to remind ourselves how often we do, as a matter of fact, both make and accept demands to justify things that we have not deliberately chosen, and which may not be under our direct voluntary control. Our lives are pervaded by the making and receiving of such demands: ''How could you have forgotten our anniversary?'' ''Didn't you realize how insensitive that was?'' ''How could you have left our child unattended at the park?'' ''What in the world do you find funny about that?'' ''Why are you so angry over such a minor slight?'' ''How can you, of all people, be so unforgiving?'' These sorts of demands are not rare or exceptional; they are the very stuff of moral life.

I think those of us who write on questions of moral responsibility often overlook just how common these sorts of demands are. We tend to focus on hard cases involving bizarre causal interventions by evil neuroscientists or severe volitional impairments caused by drugs or mental illnesses of various sorts. These are important cases to consider, but they should not blind us to the fact that in our day-to-day lives we simply take for granted that people are responsible and answerable for much more than what they voluntarily choose to do. Consider, in this connection, the wonderful passage from E. M. Forester's story *Howard's End*, when Margaret is trying to get her bull-headed husband Henry to recognize the hypocrisy involved in his cruel and unforgiving response to her sister's adultery:

> ''Not any more of this!'' she cried. ''You shall see the connection if it kills you, Henry! You have had a mistress—I forgave you. My sister has a lover—you drive her from the house. Do you see the connection? Stupid, hypocritical, cruel—oh, contemptible!—a man who insults his wife when she's alive and cants with her memory when she's dead. A man who ruins a woman for his pleasure, and casts her off to ruin other men. And gives bad financial advice, and then says he is not responsible. These, man, are you. You can't recognize them because you cannot connect'' (Forester (1910; 1986), pp. 243–244).

---

[16] Wallace (1996) makes a strong case for the claim that moral demands can only be directed at states that can be directly controlled by the reasons expressed in moral principles, and that this condition is not met when it comes to particular states of emotion or feeling. See pp. 131–132. As I suggest in the text, I think the requirement that an agent have ''direct control'' over anything legitimately subject to moral demands is too strong

[17] For an excellent discussion of these issues, see Sher (2005).

Margaret is not blaming Henry for making bad choices; she is blaming him for a more basic failure of judgment—a failure to ‘‘connect,’’ as she says. This is a failure to *recognize* and *appreciate* his shared humanity with others, a failure to *notice* that he expects things from others that he does not demand of himself, and a failure to *acknowledge* the deep similarities between his own and others’ ethical failings. Margaret’s criticism, and the moral demand implicit in it, seems perfectly legitimate, even though Henry’s ‘‘failures’’ in this case do not seem to be failures of choice.[18]

This point about how often we do, in fact, make and receive demands to justify non-voluntary attitudes does not by itself show that the rational relations view is correct in making judgment-dependence, rather than choice, the basic condition of moral responsibility. For a volitionalist will likely argue that in these sorts of cases the criticism is appropriate only if the failures in question can be traced back to a faulty choice of some sort.[19] But when we think more carefully about the sort of demand being expressed by the criticism in these cases, I think it is clear that this demand is not directed at a person’s voluntary choices, but at her rational activity more generally. When we criticize a person for being unforgiving, for example, it seems to me we are responding to something implicit in the attitude itself, not to facts about its possible origin in a person’s prior voluntary choices. More specifically, we are responding to certain *judgments of the person* that we take to be implicit in that attitude, judgments for which we consider her to be morally answerable. The hypocrite who fails to forgive others for transgressions for which she herself has often been forgiven manifests an objectionable assessment of her own moral standing and of what it is reasonable to expect of others. These are not failures of choice, but failures of judgment, and the ‘‘demand’’ we make of her in this case is that she acknowledge the inappropriateness of her attitudes and reassess the objectionable judgments they reflect. This sort of demand does not seem unreasonable, even if the person’s failure was not a failure of choice, and even if she is not in a position to change her attitude ‘‘at will.’’

For these reasons, I do not think legitimate moral criticism presupposes that an agent has voluntarily chosen (or has voluntary control over) that for which she is criticized. What matters, rather, is whether the thing for which an agent is criticized reflects her judgments, because the justificatory demand implicit in moral criticism is a demand to reassess, modify, and in some cases apologize for those judgments. At this point, however, it might be objected that this account leaves out the distinctive force of judgments of moral blame. While it may be true that we can *rationally* criticize someone for any state that reflects a mistaken judgment of some sort, it might be argued that we can only *morally blame* someone for a mistaken choice or decision. R. Jay Wallace raises precisely this objection against Scanlon's claim that what distinguishes moral appraisal from forms of unwelcome description is that it is directed at features of a person that are in principle sensitive to that person's judgments about reasons:

---

[18] Adams (1985) calls failures of this sort ‘‘cognitive sins,’’ and discusses a number of helpful examples of such failures, pp. 17–21.

[19] Indeed, this is how Levy responds to an example I gave about forgetting a friend’s birthday: ‘‘[I]f there was nothing she might reasonably have been expected to do, which would have made her recalling her friend’s birthday more probable, then she [is not responsible]’’ Levy (2005), pp. 12–13. In a footnote he acknowledges that the thing I might reasonably have been expected to do cannot be ‘‘to have remembered my friend’s birthday,’’ for that is precisely the claim that is in dispute here. But it seems to me that this is precisely what the demand comes to in this case: As a friend, it is reasonable to expect me to remember my friends’ birthdays (and not simply to do things to make it more probable that I will remember my friends’ birthdays). And, in any event, there seems to be an infinite regress looming in this direction: what if I fail to remember to do the things that make it more probable that I will remember my friends’ birthdays? Am I responsible for this only if there were things I might reasonably have been expected to do which would have made it more probable that I would recall to do the things necessary to make it more probable that I will recall my friends’ birthdays? And, if so, where does this regress end?

> Note, for instance, that beliefs are like intentions in their being explicable by reference to reasons. Thus we may criticize peoples' political and aesthetic and scientific opinions if they do not seem to us to be well justified....But criticism of a person's opinions in terms of reasons is normally very different from moral blame for a person's actions and decisions. Indeed it is different precisely in that it lacks the distinctive force of judgments of moral blame and moral blameworthiness, the connection to attitudes that gives those judgments their special 'depth.' It is one thing to criticize a philosopher's views about causation and quite another thing to blame the philosopher for supporting racist or sexist hiring practices; the difference seems to consist in the fact that moral blame has a quality of opprobrium that is lacking in criticism of beliefs or opinions. Hence we cannot hope to account for this special force solely in terms of the connection of moral blame and responsibility with justification (Wallace, 1996, pp. 80–81).

In this passage, Wallace makes the correct point that *moral* criticism directed at an action or decision normally has a kind of significance that *non-moral* criticism directed at a belief or opinion does not, but appears to draw from it the incorrect conclusion that we *cannot* be subject to moral criticism for our beliefs and opinions (''moral blame has a quality of opprobrium that is lacking in criticism of beliefs and opinions''). But this seems to conflate two different distinctions. What makes the criticism of the philosopher's action, but not his belief, *morally* significant in the case he describes is simply that this particular action is morally objectionable while this particular belief is not.[20] If the philosopher's views were not about causation, however, but about the inherent inferiority of blacks and women, I think our criticism would have quite the same ''quality of opprobrium'' that attaches to our criticism of his action. Likewise, if his action were not in support of racist or sexist hiring practices, but in support of research into supernatural causation, our criticism would lack the quality of opprobrium he describes, and would have a different kind of significance (perhaps we will be reluctant to hire him to our faculty, for example). So we must distinguish between the *depth* of a form of appraisal, which concerns whether the person can legitimately be asked to justify that for which he is being appraised, and the *significance* of it, which concerns the kind of importance that attaches to his failing to meet the normative standards that apply in the particular case.

This challenge helps to bring out three important features of this account of moral criticism that deserve comment and warrant further discussion. First, what makes moral criticism *deep*, on this view, is not its particular *content* but the fact that it involves (implicit) demands for reasons or justification. Since it is only appropriate to demand justification for things that are sensitive to a person's judgments about reasons, this reactive entitlement determines the general range of things that can serve as legitimate *objects* of moral assessment. Our voluntary actions certainly fall into to this class, but so too do our desires, emotions, and beliefs. Second, what makes moral criticism especially *significant*, on this view, is not (simply) its *depth* but its particular *content*. Because we normally take failures to respond to moral demands to involve violations of what we owe to other people, the kinds of criticisms we are open to for such failures tend to have a special kind of importance for our relations with them. Failures to respond to the normative requirements in other domains (e.g., aesthetic or epistemic) will have different kinds of significance, depending upon our reasons for caring about the demands in question.[21] Third, because the *depth* of a form of assessment is not a function of its content, but of whether it involves (implicit) demands for reasons, certain forms of non-moral assessment will also count as "deep" in the relevant sense. For example, we may criticize someone for having insufficiently justified philosophical beliefs or uninformed opinions about the musical value of hip-hop without intending these as moral criticisms. What determines the depth of a particular form of assessment depends on whether it is directed at something that reflects a person's judgment; what determines the significance of a particular form of assessment depends on the particular standards upon which the demands for reasons are based.

---

[20] It could be that Wallace is working in this passage with an implicit assumption about the substantive content of morality—namely, that moral demands apply only to intentions or actions, and therefore that we cannot be open to moral criticism for our beliefs and other attitudes. But this would beg the question against Scanlon, and it is not supported by anything he says in this passage. He does, however, argue for this position later on in his book, where he claims that moral obligations apply only to choices expressed in action. See Wallace (1996), pp. 127–136.

I am emphasizing this distinction between the depth of a form of assessment and its significance in order to make clear that the conditions of moral responsibility correlate with depth rather than significance. I am just as *morally responsible*, and just as deeply morally responsible, for an imprudent action as I am for an evil one. What makes me morally responsible in both cases is that these actions reflect my assessment of reasons, and therefore I can, in principle, be called upon to defend them and am open to rational (and in some cases moral) criticism if an adequate defense cannot be provided. Of course, since imprudence is not normally regarded as a violation of *moral* demands, the kind of criticism to which I am open in this case will not be moral criticism, and there may be no one other than myself who is in a position to legitimately make such criticisms of me.[22] Philosophers sometimes mark this difference between the two cases by saying that I am "morally responsible" for my evil action, but not for my imprudent one. This is fine, so long as we keep in mind that all that is meant by saying that someone is "morally responsible" for something is that (1) she is responsible for it (in the sense that it is attributable to her in a way that makes moral appraisal, in principle, appropriate), and (2) because of the particular content of the action or attitude in question—i.e., the fact that it involves violations of what we owe to other people—the kind of appraisal that is appropriate in this case is moral in nature. I am inclined, myself, to say that we are ‘‘morally responsible’’ for both actions (since both are attributable to us in the relevant sense), but I can see that this may have infelicitous connotations of its own, insofar as it suggests that we can be morally criticized for imprudence. But this terminological dispute is not important as long as we are clear about the difference between the depth of a form of assessment (which concerns whether the person is responsible for the thing in question) and its significance (which concerns the nature of the normative failing in question).

---

[21] When I talk of the different kinds of "significance" that different sorts of normative failings can have, I do not mean to be implying that moral failings are always more serious than non-moral failings. Some moral failings can be quite trivial (e.g., taking an extra item through the express check-out line at the grocery store), while some non-moral failings can be catastrophic (e.g., an engineering miscalculation which results in the deaths of hundreds of people). My point is only that moral failings involve breaches (sometimes trivial) of what we owe to other people, and therefore they have direct implications for our relations with them.

[22] But what would make self-*criticism*, rather than mere negative self-*evaluation*, appropriate here is that what is being assessed is something which is, in principle, sensitive to my judgments about reasons, and for which I can therefore reproach myself if I decide that those judgments have been faulty. Contrast this to the negative self-evaluation one might make of oneself on the basis of one's appearance or lack of hand-eye coordination.

But does the fact that a form of assessment involves this (in principle) demand for reasons capture all of the features of ‘‘deep’’ assessment that we identified in discussing the views of Wolf and Smart above? Deep assessment, recall, is supposed to involve (1) ‘‘judging’’ not simply ‘‘describing’’ or ‘‘grading’’ (2) ascriptions of credit or fault, and (3) a reactive entitlement which includes certain attitudes and emotional responses (e.g., indignation, resentment, and esteem). I believe the account just given captures all of these features. Deep assessment, of which moral appraisal is just one form, makes claims about a person’s judgmental activity, about the way she recognizes, assesses, and responds to reasons. Unlike a person’s physical traits or innate talents, this is something for which it makes sense to hold a person responsible. If a person’s judgments, as manifested in her actions and attitudes, appear to violate certain normative standards (whether those be moral, philosophical, prudential, or whatever), it is appropriate (in principle) to ask her to reassess those judgments and to explain, justify, modify, and in some cases apologize for her actions or attitudes in light of this reassessment. Criticism, in this case, is not mere unwelcome description, but calls upon a person to re-evaluate the grounds of her attitudes and intentions and to modify them if those grounds seem faulty or insufficient.[23] This kind of criticism, therefore, would not be appropriately directed at a feature that is not expressive of a person’s judgment. Finally, depending upon the particular normative demands in question, this kind of criticism supports certain kinds of reactions, including hurt feelings, indignation, resentment, distrust, and so on.

The difference between moral appraisal and other forms of deep assessment will come out most clearly in the kinds of reactions we normally have to moral failures as distinct from failures of other sorts. Since morally objectionable actions and attitudes involve violations of what we owe to other people, they seem to pose a direct challenge to the moral standing and value of those to whom they are directed. It is reasonable for those wronged in such cases to respond to such a challenge with resentment and hurt feelings, and for third parties to respond with indignation. The special significance of moral as opposed to other forms of deep criticism, then, is that it concerns the quality of our relations with others rather than the quality of our activity in some other normative domain (such as philosophy, or music). To fail to recognize or care about reasons flowing from the value of other people has direct implications for our relations with them, in a way that failing to recognize or care about reasons flowing from the value of philosophy, or jazz, or basketball does not.[24] Although I am open to criticism if my philosophical arguments lack rigor or consistency, and although this criticism, if correct, is a serious one for someone trying to pursue a career in this discipline, such criticism does not concern my attitudes toward other people and therefore does not call into question my relations with them. This is why philosophical and other forms of non-moral criticism (e.g., aesthetic, prudential, epistemic) normally do not support reactive attitudes on the part of others such as hurt feelings, resentment, or indignation.

---

[23] It might be thought here that a demand to ‘‘modify’’ one’s attitudes implies that one must have volitional control over them. But this is a mistake. We cannot modify our attitudes ‘‘at will,’’ though we can re-evaluate the grounds upon which they are held and we may come to see that those grounds are mistaken. In making such an assessment, our attitudes will usually change; but we have not changed them ‘‘at will.’’ For a very helpful discussion of these two different senses of ‘‘control’’ we might exercise with respect to our attitudes, see Hieronymi (2006). For related discussion of the voluntary/involuntary distinction, see Adams (1985), pp. 6–11.

I hope to have shown in this section that the rational relations account of moral responsibility can license precisely the kinds of "deep" moral assessments that volitionalist accounts purport to justify. Once we see what it is that provides the "depth" to these forms of assessment (that they concern a person's activity in some area, and that they address a justificatory demand to their target), I do not see why such appraisals should be limited to those things a person has voluntarily chosen rather than to those things that reflect her rational judgments more generally. What this means, however, is that there is very little room on the rational relations view to say that a person has "bad attitudes" but that she is not responsible, and therefore not blameworthy, for those attitudes. For to say that a person's attitude is "bad," on this account, is to say that she has *judged badly*, which implies responsibility and culpability on the part of the person being assessed. It might be thought that this failure to distinguish between bad agents and blameworthy agents counts against the rational relations view. In the next section, however, I will try to show that the limited relevance of this distinction on this view is in fact one of the strongest marks in its favor.

5 Bad agents and blameworthy agents

Even if I am correct in claiming that the vast majority of our desires, emotions, and other attitudes reflect our judgments or evaluative commitments, it might still seem that there is a big difference between saying that we are open to negative assessment on the basis of these attitudes and saying that we are blameworthy for them. To many, it just does not make sense to say that a person can be *blameworthy* for an attitude she has not chosen and that is not under her voluntary control. Neil Levy states this objection nicely when he writes:

> Volitionists agree that we can assess agents upon the basis of their morally relevant attitudes, as the attributionists claim. What they deny is that finding that an agent is morally flawed is necessarily to hold that agent responsible for

---

[24] Of course, the values we do and do not care about will have indirect implications for our relations with others, in the sense that they will help determine with whom we want to spend time, and what shared activities are open to us in these relations. And being in close personal relationships can give us moral reasons to take seriously and to try to appreciate the values which our loved ones care about. Failing to make such efforts can be grounds for moral criticism in such cases, but not because we have failed to respond appropriately to the values in question, but because we have failed to respond appropriately to our loved ones.

> her flaws; that all negative assessment is blame. After all, volitionists might point out, even attributionists find fault with agents without holding them to be at fault: when these faults are the product of transient mental illness, for instance. It is, therefore, false that there is no conceptual room for a distinction between the bad and the blameworthy, and attributionists owe us an argument for closing the gap between the two (Levy, 2005, p. 6).

I agree that there is a general distinction to be drawn between ''the bad'' and ''the blameworthy'' when it comes to assessing features of persons. We can all acknowledge that faulty hearing or general clumsiness is a bad thing without thereby claiming that a person who possesses such traits is blameworthy for them. And such a distinction is certainly relevant when it comes to the assessment of non-rational animals and young children. I can assess a dog as ''vicious'' and a toddler as ''bratty'' without claiming that they are responsible and therefore blameworthy for having these traits. When we call a dog ''vicious,'' we imply that it has a ''bad disposition,'' and that we should steer clear of it if we care about our physical well-being. But we do not think that a dog's disposition to attack us reflects a judgment that we lack moral value or standing, or indeed that it reflects any judgment at all on the dog's part about what reasons it has. Therefore, it would make no sense to say that the dog is ''blameworthy'' or ''at fault'' for his vicious disposition, because moral criticism, as I have argued, involves an implicit demand for justification; it makes no sense to make that demand of a creature who cannot recognize, assess, and respond to reasons. The same thing can be said of young children whose attitudes and behavior are not yet reasons-responsive, and human beings who are subject to ''transient mental illnesses'': if a paranoid schizophrenic attacks me in a moment of severe mental delusion, I may see that action or his condition as ''bad'' but I will not consider him blameworthy for it (indeed, I will not assess him or his behavior in moral terms at all). Why? Because such an act cannot be said to reflect his judgments, and therefore it would make no sense to demand that he justify it. But if I assess an (otherwise normal) adult human being as ''cruel'' I am not merely attributing a trait to him for which he may or may not be responsible (as when I say that a dog has a ''vicious disposition''). I am (in part) making a demand of him, a demand that he justify the objectionable judgments his actions and attitudes express concerning the moral status of others. This demand *by its very nature* implies responsibility, for it is directed at his judgmental activity, activity for which we must regard him as responsible if we are to regard him as a moral agent in any sense. To say that a person's judgmental activity is bad but that he is not responsible for it is, in effect, to say that he is not to be regarded as someone to be reasoned with, but merely as someone to be understood, treated, managed, or controlled.[25] It is to regard a person as we would regard a vicious dog or a bratty toddler, someone to be avoided and/or trained, if possible, but not someone with whom it is possible to enter into relationships of mutual respect and recognition. The distinction between ''bad agents'' and ''blameworthy agents'' is, therefore, a distinction we should not be anxious to invoke except in the very rarest of circumstances. And it should not count as a strike against non-volitional accounts that they leave so very little space for drawing this distinction in our moral assessment of persons.[26]

---

[25] Here, of course, I am drawing upon Peter Strawson's influential discussion of the difference between taking up a ''participant stance'' and an ''objective stance'' toward an agent. See Strawson (1993), p. 59.

Levy might well respond that I am missing the essential point, which is that some people are responsible for *becoming* cruel (e.g., by deliberately pursuing courses of action that would harden them to the claims of others), while other people acquired their cruel attitudes through no fault of their own (e.g., in virtue of having had an abusive upbringing). This is the intuition that makes us uneasy when we consider cases like that of Robert Alton Harris, who carried out horrifically cruel actions and yet was also the victim of horrific abuse as a child.[27] In such cases, there is clearly a temptation to say, as Levy suggests, that Harris was "morally flawed" but that he was not responsible for his flaws. Yet when we consider more carefully the implications of that claim, I think it becomes considerably less appealing. For the "moral flaw" in this case was Harris's own judgmental activity, his own evaluation of the weight and significance of the claims presented by others.[28] Thus Levy's view would commit us to denying Harris responsibility for his own judgments—literally, for what he thinks—which is tantamount to denying him basic status as a moral agent. Harris himself pointedly rejected this denial of his own moral agency.[29]

In thinking about cases of this sort, then, I think we would do well to distinguish two different questions: the question of one's responsibility for becoming a certain sort of person, and the question of one's responsibility for the judgments expressed in one's actions and attitudes. It will be a complex story, for each and every one of us, how we became the sorts of people we are, with the particular values, interests, cares and concerns that we hold; and very few, if any, of us can plausibly claim to bear full or even substantial responsibility for how we became the particular people we are. Even so, I submit, we cannot help but regard ourselves as responsible and answerable for the particular judgments expressed in our actions and attitudes, regardless of what circumstances may have shaped these assessments. A cruel person, whether she became so through her own efforts or through unfortunate circumstances, is someone who judges that the fact that something will cause pain or suffering to another is no reason to avoid it (and is perhaps even a reason to pursue it). It is that judgment, as reflected in her actions and attitudes, for which we consider her answerable and expect her to defend or give up.[30] Whether she is also responsible for becoming the sort of person who makes such judgments is a different question and may well have a different answer.

---

[26] It might be thought that non-volitionalist accounts do not leave any space at all for drawing this distinction. But I think even a non-volitionalist can allow for such a distinction in the sorts of fanciful cases involving "implanted" or hypnotically-induced attitudes. For example, we can imagine a person induced through hypnosis to think "Jones is evil" every time the phone rings, even though Jones is his best friend and he is puzzled by the thought whenever it occurs. In this case, we might say the attitude is "bad" but he is not blameworthy for it. But this is because the "attitude" (if we can call it that) in this case does not really reflect his own evaluative assessment of reasons. It would make no sense to ask him to defend it, because he himself regards it as completely puzzling and unintelligible whenever it occurs.

[27] Watson (1993) devotes an extended discussion to this case in his thought-provoking article "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." I am much indebted to his discussion.

[28] After shooting two young men in cold blood, Harris reportedly suggested to his accomplice that it would be amusing if they dressed up as police officers to inform their parents that their sons had been killed (Watson, 1993, p. 132). This is someone who clearly was able to make judgments about reasons (e.g., about what sorts of things would be "fun" or "amusing"), but these judgments gave no weight or significance at all to the lives or well-being of others.

[29] According to his sister, "He told me he had his chance, he took the road to hell and there's nothing more to say" (Watson, 1993, p. 133).

Having said that, however, it is also important to point out that these facts about a person's formative circumstances can make a big difference when it comes to questions of moral assessment (as distinct from questions of responsibility). There is room on the rational relations view to say that a person who was the victim of unfortunate formative circumstances may be open to less serious moral criticism for her vicious attitudes and actions than a person manifesting similar attitudes and actions who did not face such challenging circumstances.[31] For one of the things we can take into account when assessing a person's attitudes and actions is how difficult it may be for her, given the brutal conditions of her upbringing, to appreciate the weight and significance of the reasons presented by the lives and interests of others. We need not say that a person is *not responsible* for her judgments in order to take into account the relevance of her formative circumstances when it comes to our moral assessment of her in light of those judgments. We can acknowledge that a person who has not been exposed to morally appropriate values may have a more difficult time appreciating the importance of these values than someone who has had the benefit of a decent moral upbringing. Such people may be open to less serious moral criticism for their vicious attitudes than others, but they are just as fully responsible for them.

For these reasons, I think it a mark in favor of the rational relations view that it makes it so difficult to draw the distinction between ‘‘morally bad'' agents and ‘‘morally blameworthy'' agents. This distinction itself would require us to regard some agents as the passive victims of their faulty judgments, as I was the passive victim of my faulty hearing. I think this is a dangerously patronizing and disrespectful stance to take toward another human being, one that we should be very reluctant to resort to in practice.

---

[30] Notice here that in calling someone ‘‘cruel,'' we presuppose that she understands that her actions or attitudes cause pain or suffering to others; the ‘‘cruelty'' comes from the fact that she does not regard this as a reason to act differently (or that she sees it as a positive reason to continue acting as she does). In a very puzzling section of his paper, Levy claims that attributionists in general, and Scanlon in particular, are committed to claiming that agents are blameworthy for actions that cause harm to others even if they had no way of knowing that their actions were causing harm (e.g., if, unbeknownst to us, plants could feel pain, we would be blameworthy for treading on them) (Levy, 2005, p. 9). After all, such harmful acts are ‘‘attributable'' to them in the relevant sense. But this is a mistake. The attributionist, like the volitionist, can say that an agent is blameworthy for an action or attitude only if that action or attitude violates a moral norm we expect reasonable persons to accept. Since reasonable persons have no reason to think that plants can feel pain, a person whose attitudes and actions reflect this judgment is not ‘‘blameworthy'' for these things. I think Levy here is conflating questions of responsibility and questions of blameworthiness. Non-volitionists, no less than volitionists, can acknowledge the existence of epistemic conditions when it comes to determining whether a person is blameworthy for her actions and attitudes; these conditions do not seem to be relevant to determining whether a person is responsible for her actions and attitudes, however.

[31] I discuss this point at greater length in Smith (2005), pp. 267–270.

6 Conclusion

My aim in this paper has been to respond to a worry that might be raised against some recent non-volitionalist accounts of moral responsibility: namely, that these accounts do not provide a genuine alternative to traditional volitional accounts of responsibility, because the conditions of responsibility they offer do not license the sorts of "deep" moral assessments involved in our current practices of moral responsibility. In response, I have tried to give a fuller account of the nature of moral appraisal and how it differs from certain other forms of assessment, and I have tried to show why my own non-volitional account supports robust forms of moral criticism. I have also conceded that this view does not leave much space for the distinction between "bad agents" and "blameworthy agents," because to say that an attitude or an action is "morally bad" on this view is to say that an agent has *judged badly*, which is an assessment that implies both responsibility and blameworthiness. However, I have tried to show that this is a distinction we should not be eager to place weight on in our moral practices anyway, for to regard a person as "morally bad" while refusing to see her as "morally blameworthy" is to deny her basic standing as a moral agent. At the end of the day, I submit, it is much more respectful to be blamed for our moral faults than to be pitied for them.

Acknowledgements I would like to thank Christine Korsgaard, Richard Moran, and T. M. Scanlon for very helpful comments on an earlier draft of this paper. I am also grateful to Rahul Kumar and Sandra Reiter, who provided invaluable feedback on more recent drafts of this material.

References

Adams, R. (1985). Involuntary sins. *The Philosophical Review, 93*(1), 3–31.
Blum, L. (1980). *Friendship, altruism, and morality*. London: Routledge & Kegan Paul.
Dennett, D. (1984). *Elbow room: the varieties of free will worth wanting*. Cambridge, MA: MIT Press.
Fischer, J. M., & Ravizza, M. (Eds.) (1993). *Perspectives on moral responsibility*. Ithaca: Cornell University Press.
Forester, E. M. (1910; 1986). *Howard's end*. New York: Penguin Books, Inc.
Hieronymi, P. (2006). Controlling attitudes. *Pacific Philosophical Quarterly, 87*, 45–74.
Hume, D. (1751). *Enquiry concerning the principles of morals*. In L. A. Selby-Bigge & P. H. Nidditch (Eds.), (1975) *David hume: enquiries concerning human understanding and concerning the principles of morals*. 3rd ed. Oxford: Clarendon Press.
Hume, D. (1739). *A treatise of human nature*. In L. A. Selby-Bigge, & P. H. Nidditch (Eds.), (1978) *A treatise of human nature*, 2nd ed. Oxford: Oxford University Press.
Levy, N. (2005). The good the bad and the blameworthy. *Journal of Ethics and Social Philosophy, 1*(2), 1–16.
Oakley, J. (1992). *Morality and the emotions*. London: Routledge.
Sankowski, E. (1977). Responsibility of persons for their emotions. *Canadian Journal of Philosophy , 8*(4), 829–840.
Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
Scanlon, T. M. (1988). The significance of choice. In S. McMurrin (Ed.), *The Tanner Lectures on Human Values*, vol. 8. (pp. 149–216) Salt Lake City: University of Utah Press.
Schlick, M. (1939). When is a man responsible? D. Rynin (Trans.) *Problems of ethics*. New York: Prentice Hall, Inc., 143–158.
Sher, G. (2005). Kantian fairness. *Philosophical Issues, 15*, 179–192.
Sher, G. (2006). Out of control. *Ethics, 116*(2), 285–301.
Slote, M. (1992). *From morality to virtue*. Oxford: Oxford University Press.
Smart, J. J. C. (1961). Free-will, praise, and blame. *Mind, 70*(279), 291–306.

Smith, A. (2000). Identification and responsibility. In A. van den Beld (Ed.), *Moral responsibility and ontology* (pp. 233–246). Dordrecht: Kluwer Academic Publications.

Smith, A. (2004). Conflicting attitudes, moral agency, and conceptions of the self. *Philosophical Topics*, *32*(1 & 2), 331–352.

Smith, A. (2005). Responsibility for attitudes: activity and passivity in mental life. *Ethics*, *115*(2), 236–271.

Smith, A. (2007). Forthcoming. On being responsible and holding responsible. *Journal of Ethics*.

Strawson, P. (1993). Freedom and resentment. In J. M. Fischer & M. Ravizza (Eds.), *Perspectives on moral responsibility*, (pp. 45–66). Ithaca: Cornell University Press.

van Inwagen, P. (1980). *An essay on free will*, Oxford: Oxford University Press.

Wallace, R. J. (1996). *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.

Watson, G. (1996). Two faces of responsibility, *Philosophical Topics, 24*(2), 227–248.

Watson, G. (1993). Responsibility and the limits of evil: Variations on a Strawsonian theme. In J. M. Fischer & M. Ravizza (Eds.), *Perspectives on moral responsibility* (pp. 119–148). Ithaca: Cornell University Press.

Wolf, S. (1990). *Freedom within reason*. Oxford: Oxford University Press.