https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

QUALITIES OF WILL*

By David Shoemaker

> The reactive attitudes . . . are essentially reactions to the quality of others' wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern.[1]
>
> —P. F. Strawson

What unites disparate theorists of moral responsibility is that they all seek to identify the nontrivial conditions for being morally responsible for something, most typically understood as being the worthy target of a range of responses like praise or blame. Many such theorists think we need to appeal to a wide variety of agential or non-agential features, ultimately requiring a complicated conjunction of conditions, including authenticity, autonomy, historical, control, and/or epistemic conditions for responsibility. The resulting theories can, at the least, be quite cumbersome. P. F. Strawson suggested an elegant alternative, one that attempted to account for the entirety of our range of "responsibility responses" exclusively in terms of the targeted agent's *quality of will.* Gary Watson presented a powerful objection to this approach, but in recent years several new versions have cropped up that might seem to avoid the objection. My aim in this essay is twofold. First, I intend to show that all of these more refined quality of will theories are still subject to an expanded version of Watson's objection, one drawn from our responses to real life, "marginal" cases. Second, I will show that the general quality of will

* For providing very helpful remarks on an earlier draft of this essay, I thank Angela Smith and Matt Talbert. For their insightful questions and comments at my presentations of this material, I am grateful to several contributors to this volume (especially David Brink and Erin Kelly), Victor Kumar, David Lefkowitz, Eugene Mills, and Nancy Schauber. I am also grateful for the financial support of the Freedom Center, Tulane University, and a state of Louisiana ATLAS Grant while I worked on the paper. For their discussion and help with various aspects of the project, I thank Justin D'Arms, Terry Horgan, Dan Jacobson, Shaun Nichols, Connie Rosati, and Mark Timmons. For many enjoyable and exceedingly fruitful conversations about this material, in particular regarding the distinction between answerability and accountability, I am grateful to Steve Wall. My largest debt, though, goes to Michael McKenna, who, through detailed written comments, excellent questions, and patient (and lengthy) discussion, made me see many of my earlier errors but also helped me to express many of the ideas here in a much clearer and defensible fashion. Any mistakes and unclarities that remain, however, are most assuredly attributable to me, and I remain answerable for them.

[1] P. F. Strawson, "Freedom and Resentment," in *Free Will,* 2d edition, ed. Gary Watson (Oxford: Oxford University Press, 2003), 83.

doi:10.1017/S0265052513000058
© 2013 Social Philosophy & Policy Foundation. Printed in the USA.

approach is still salvageable—and may in fact be very plausible—but only if it embraces a *pluralistic* approach to responsibility.

I. Pure Quality of Will

Call a *Pure Quality of Will* (PQW) theory of responsibility one that puts forward agents' quality of will as exclusively filling two central roles: (a) to serve as the target of all our responsibility responses, and so (b) to serve as the unifying explanation of excuses and exemptions from responsibility. Regarding (a), as Fischer and Ravizza put it, there is a "spectrum of feelings, attitudes, and practices" that are "appropriately applied only to other persons," in virtue of their being "morally responsible for what they do."[2] In other words, there is a roughly unified and unique set of responses we have to agents that we think of as implicating their responsibility. These "responsibility responses" are the ways in which we may be said to *hold* agents responsible. They include, of course, Strawson's famous core reactive attitudes—resentment, indignation, and guilt—but they also include, more generally, practices of praise, blame, reward, and sanction.[3] On a PQW theory, when we hold someone responsible via one or more of these responses we hold her responsible exclusively for her quality of will.

Regarding (b), if what we are exclusively held responsible for is our quality of will, then something about it should explain why we occasionally get off the hook, that is, explain why we are sometimes *not* subject to the responsibility responses for what we do. Excusing pleas are those that, when followed, get one off the hook only for a specific action or attitude, and they include "It was an accident," or "I could not help it" (where this is buttressed by something like "I was pushed," or "It was the only way"). Exemptions, on the other hand, get one off the hook qua agent: one is not the kind of creature to whom responsibility responses coherently or appropriately apply. Exempting pleas include "He's just not himself," or "He's only a child," or "He's a hopeless schizophrenic."[4] So if responsibility is exclusively for one's quality of will, one's not being responsible (in being excused or exempted) must involve some story about one's quality of will not being implicated.

[2] John Martin Fischer and Mark Ravizza, "Introduction," in *Perspectives on Moral Responsibility* (Ithaca, NY: Cornell University Press, 1993), 6. This spectrum marks what they call our "responsibility-reactions." I call them our "responsibility responses" not only because I think the phrase is a bit snappier, but also because "response" better connotes the *engagement* with the target that is indicative of many such responses.

[3] See also, John Martin Fischer, "Introduction," in *Moral Responsibility*, (Ithaca, NY: Cornell University Press, 1986), 12.

[4] Strawson, 77–79. I here follow Gary Watson's labeling of "excuses" and "exemptions" in "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," in Gary Watson, *Agency and Answerability* (Oxford: Oxford University Press, 2004), 219–59.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

For Strawson, the advantages of a successful PQW theory would be enormous, for it could purportedly account for all of the deeply human, nonconsequentialist features of our practices of responsibility without having to go beyond the facts as we know them to justify these practices via the "panicky metaphysics of libertarianism."[5] Our practices have distinctive features that are richly interpersonal and are about far more than mere social regulation; they express our very humanity in occasionally retributive ways. If what generates this robust range of responses, however, is just the actual quality of the targeted agent's will, then we should have no need to worry what some theoretical fact about the truth or falsity of determinism might imply about the *freedom* of that will, about alternative wills one could have had, or about whether one was the ultimate source of that will.

This is a terribly ambitious argument, and many have found fault with some aspect of it or another. In particular, several theorists have pounced on the fact that various of Strawson's own excuses or exemptions may be generalized in ways that bring worries about determinism back to the fore, or even that quality of will itself may have to include a free will or free action component in order to explain our sensitivities to it.[6] Consequently, I here want to set aside the determinism/freedom aspects of Strawson's argument in order to focus just on the prospects for a PQW approach with respect to the two features previously mentioned. The relevant question, then, is whether there could be such a theory that genuinely accounts for *all* of our responsibility responses, in both their deployment and their suspension.

In objecting to this possibility as formulated by Strawson, Gary Watson noted, "A child can be malicious, a psychotic can be hostile, a sociopath indifferent, a person under great strain can be rude . . . ."[7] Watson's point was that if quality of will were all that mattered for responsibility, we could not explain why certain negative reactive attitudes like resentment and indignation toward such agents are appropriately suspended, since these individuals do seem to have a capacity for exhibiting objectionable quality of will in *some* sense. Watson takes this to be prima facie evidence that "reactive attitudes are sensitive not only to the quality of others' wills, but depend as well upon a background of beliefs about the objects

[5] Strawson, 93.

[6] See, e.g., Watson, in "Responsibility and the Limits of Evil"; David Wiggins, "Towards a Reasonable Libertarianism," in Watson, *Free Will,* 94–121; Paul Russell, "Strawson's Way of Naturalizing Responsibility," *Ethics* 102 (1992): 287–302; R. J. Wallace, *Freedom and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994); and Michael McKenna, "Where Frankfurt and Strawson Meet," *Midwest Studies in Philosophy* 29 (2005): 163–80.

[7] Watson, *Agency and Answerability,* 228. I have left out Watson's mention of someone unfortunate in formative circumstances who is nevertheless "cruel," as this case is a very difficult one, and proper treatment of it would take us too far afield. For an insightful discussion of the complexities, see Michael McKenna, "The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism," *The Journal of Ethics* 2 (1998): 123–42.

of those attitudes."[8] Our theory of exemptions, thus must incorporate these additional facts (likely facts about history, control, epistemic capacities, and so on), but "[t]o the extent that some such account seems necessary, [the enterprise of the PQW theorist] is doomed."[9]

Nevertheless, several recent PQW theories seem designed to avoid this worry. In the next section, I briefly survey them.

II. Three Qualities of Will

An initial and obvious problem with constructing a PQW theory is that it is just not clear what a quality of will *is.* Strawson's own gloss in the epigraph does not help, for the question is what is a good or ill *will,* precisely? There have actually been three competing understandings of "quality of will" offered in the literature, generating three competing PQW theories of responsibility (some explicitly non-Strawsonian). The first theorist to attempt to refine the notion of "quality of will" in a way that could enable it to carry the burdens of a PQW theory was T. M. Scanlon in "The Significance of Choice."[10] He developed the version I will call the *Pure Quality of Judgment* (PQJ) view. Since his analysis is intended to dovetail with his larger moral theory, its content is explicitly contractualist: quality of will is about self-governance, which is determined by the degree to which one attends and adheres to the principles no one could reasonably reject within a system of co-deliberation. The will whose quality is to be assessed is therefore the agent's *judgment* (or lack thereof) about the relevant reasons for actions and attitudes. On Scanlon's treatment, such judgment is what is "up to us," and so we can be properly "held responsible" for our judgment-sensitive attitudes "in several central senses of that phrase: they can be properly attributed to us, and we can properly be asked to defend them—to justify the judgment they reflect."[11] Negative responsibility responses are thus warranted for judgment-sensitive attitudes in virtue of their reflecting either defective or absent judgment, what Scanlon labels "faulty self-governance."[12]

So how does the PQJ story offer a unifying explanation of exemptions and excuses that sidesteps Watson's worry? To view others as blameworthy, on this account, is to believe that there is a gap between the way they governed themselves and the relevant standards of judgment. Successful excuses, e.g., accidents, reveal that the agents governed them-

[8] Watson, *Agency and Answerability,* 228.

[9] Ibid.

[10] In Sterling McMurrin, ed., *The Tanner Lectures on Human Values,* 8 (Salt Lake City: University of Utah Press, 1988), 149–216. Wallace (124–33) offers a similar view, targeting "quality of choice." For ease of exposition, I will focus solely on Scanlon's view, although he and Wallace disagree on key points.

[11] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, Belknap Press, 1998), 22.

[12] Ibid., 268.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

selves without fault, despite initial appearances. And what of exemptions? Essential to the PQJ story is the agent's *capacity* for "critically reflective, rational self-governance."[13] If you are incapable of governing yourself via judgment, then your self-governance has *no* quality; it is not faulty, it is just nonexistent. Exemptions apply to those lacking this capacity, e.g., children and psychotics: while they may have poor quality of will in some other sense, they are incapacitated for responsibility only in virtue of their lacking the capacity for quality of judgment. Both excuses and exemptions are thus unified solely by appeal to quality of judgment, and there is ostensibly no need to look beyond such facts.

The second interpretation of "quality of will" is much more consciously Strawsonian, and it has been most fully developed recently by Michael McKenna.[14] Call it the *Pure Quality of Regard* (PQR) view. On McKenna's understanding, quality of will consists in the worth of the "regard or concern one has for others (or oneself), and toward the relevance of moral considerations, as manifested in one's conduct."[15] This regard is "*exhibited* in the reasons for which one acts, the intention with which one acts, or the choices one makes,"[16] but it may not necessarily be found *in* any of these, given that they could each be morally innocuous.[17]

So how would a PQR view theoretically explain and unify excuses and exemptions? There is a structurally similar story to be told as on the PQJ view. To believe someone is blameworthy is to believe, in part, that she has expressed insufficiently high regard for others or for moral considerations in her actions or attitudes. Successful excuses reveal that what was initially thought to be poor regard actually was not. For example, when what one did was accidental, there was no connection between one's actual regard and one's behavior. To see how exemptions might work to stave off Watson's worry in terms of quality of

[13] Scanlon, "The Significance of Choice," 174.

[14] Michael McKenna, *Conversation and Responsibility* (New York: Oxford University Press, 2012). McKenna himself does not advocate a *pure* Quality of Will theory, for one, because he admits that "attributability" is a distinct conception from "accountability." The latter is his primary focus in the book. Nevertheless, it's not clear even with respect to accountability alone that he's offering a PQW theory. That's because, while he thinks quality of will is at least *necessary* for accountability, it may not be sufficient (he remains agnostic on this point, given his uncertainty about whether, for example, freedom and epistemic conditions are incorporated into quality of will). See, for example, 61.

[15] McKenna, *Conversation and Responsibility,* 59. Nomy Arpaly has put forward a PQR view whose scalar dimension (i.e., what it is that makes actions *more* or *less* blameworthy and praiseworthy) requires only this latter feature, the concern one has for moral considerations, whether one is aware of those considerations *as* moral or not. See her *Unprincipled Virtue* (Oxford: Oxford University Press, 2003), 79–84.

[16] McKenna, *Conversation and Responsibility,* 59; emphasis mine.

[17] This is what marks one difference, I think, between McKenna and Scanlon, given that Scanlon says explicitly that one's attitude toward others is *constituted* by "[t]he agent's reasons for acting (and the fact that other considerations did not count for him as reasons against so acting). . . ." T. M. Scanlon, *Moral Dimensions* (Cambridge, MA: Harvard University Press, Belknap Press, 2008), 152–53. I will discuss what I take to be the most important difference between judgment and regard later in the paper.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

regard requires seeing the reactive attitudes as communicative, their function being to express the basic demand for good will. As Watson originally suggested, for reactive attitudes to be intelligible as forms of communication, their targets must be capable of understanding that demand. What those individuals exempted from the moral community lack are the capacities for doing so.[18] But this is just to say that facts beyond quality of will are required to explain exemptions. McKenna thus requires the agential capacity not just for understanding the basic demand but for *participating* in the practices of holding responsible. It is not just that those in the exempted class cannot understand the communication of others; it is also that they cannot communicate *themselves* within the practices. This means that if an individual is incapable of understanding the basic demand for good will as expressed via the reactive attitudes, then he is "incapable of acting from a will with a moral quality that could be a candidate for assessment from the standpoint of holding responsible."[19] Exempted agents are thus "incapable of acting from a will that *could* have the relevant kind of moral quality."[20]

A third way of understanding "quality of will" comes from Jeanette Kennett and Nicole Vincent, who, in a recent presentation, "Folk Psychology, the Reactive Attitudes, and Responsibility,"[21] introduced what I will call the *Pure Quality of Character* (PQC) view, according to which quality of will is about "character traits, personality, temperament and behavioral dispositions."[22] Importantly, Kennett and Vincent themselves do not defend a PQC theory. But one might. We do seem to have responsibility responses to people's characters (for instance, by disdaining or admiring them), so it might well be that what we are responsible for is exclusively our qualities of character, as opposed to our qualities of judgment or regard.[23] One would then be excused, on this theory, for performing actions that do not reflect a poor quality of character,[24] and one would be exempted for lacking the capacity for character in the first place. A suitably finessed theory might then explain why children and psychotics are exempt—they lack a sufficient capacity for character—as are dogs and robots.

[18] Watson, *Agency and Answerability,* e.g., 229–33.

[19] McKenna, *Conversation and Responsibility,* 78.

[20] Ibid.

[21] From the conference for the 50th anniversary of P. F. Strawson's "Freedom and Resentment," held at The College of William & Mary, September 27–28, 2012.

[22] Ibid.

[23] David Hume actually seemed to hold a view somewhere in this vicinity. See, e.g., *A Treatise of Human Nature,* ed. L. A. Selby-Bigge (Oxford: Oxford University Press, 1960), book II, part III. sec. II, 411. See also the helpful discussion in George Sher, *In Praise of Blame* (Oxford: Oxford University Press, 2006), ch. 2.

[24] So-called out of character actions might seem to pose a problem for a PQC view, but I think there are ways to respond that I will not detail here. For some possibilities, though, see Nomy Arpaly and Timothy Schroeder, "Praise, Blame, and the Whole Self," *Philosophical Studies* 93 (1999): 161–88, esp. 182–84.

III. Generalizing Watson's Objection

Each of these PQW theories purports to explain, all on its own, what we are responsible for and why some agents are therefore exempt from our responsibility responses (and from here on out I will focus only on exemptions, as they are what posed the original problem for Strawson's PQW theory). They each offer important insights. A more generalized version of Watson's worry still applies, however, because the monism of these theories prevents them from providing an adequate explanation of our *ambivalent* responsibility responses in many marginal cases. Start with a case Kennett and Vincent rightly discuss as a problem for a PQC theory: psychopathy. Psychopaths can be manipulative, selfish, greedy, cruel, and indifferent (as Watson also notes), so we may be contemptuous of or have disdain for them. They may judge that our interests are not worth respecting, so we may be disapproving or critical of them. But other sorts of response have seemed to many to be inappropriate. This is because psychopaths have significant incapacities—empathic, emotional, and/or normative—and it strikes many as unfair or pointless to *resent* or *sanction* them as a result.[25] But if this is right, then there is no single PQW story that will be satisfactory with respect to them. Both the PQC and PQJ views imply that, insofar as psychopaths have capacities for both character and rational judgment, they are simply not exempt from our responsibility responses. But they do seem exempt from *some* such responses, as we have just seen. If anything, their incapacities suggest they lack the capacity to *regard* others properly. Now this might seem to give a PQR view an edge over PQC and PQJ views, for it could explain why psychopaths are exempt from resentment and blaming sanctions in a way the latter two cannot. But then how might a PQR theorist explain why psychopaths seem to be *nonexempt* with respect to our other negative responses such as contempt, disdain, disapproval, and so forth?

There are many other marginal cases about which we are also ambivalent. The clinically depressed, for instance, may sometimes make terrible decisions about how to treat others and themselves, and we often respond to these decisions with disappointment or embarrassment while nevertheless recognizing their agents to be exempt from other sorts of negative responses (such as resentment or disdain). Or consider those with the moderate cognitive impairments of midstage Alzheimer's disease. They are surely exempt from resentment or moral indignation, but caregivers often still respond to them with annoyance or admiration. High functioning autistic people also generate ambivalent responses, often the targets of admiration, disappointment, or pride, but not resentment or indignation. And we sometimes get from caregivers the oppo-

[25] Some theorists—e.g., Matthew Talbert and Angela Smith—disagree with the latter assessment, but theirs is not a widely shared sentiment amongst responsibility theorists. At any rate, I hope to provide some reasons in favor of this assessment later in the essay.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

site sort of ambivalence to mildly mentally retarded adults, to whom resentment or indignation sometimes apply full force, while general moral disappointment or disdain do not.[26]

It may seem, then, that Watson's objection generalizes: not even these more refined pure quality of will theories can explain our ambivalent responses to marginal agents, i.e., why we seem to exempt them from some responses but not others. As a result, we may worry once more that any quality of will approach is "doomed." Such a conclusion would be too quick, however. Instead, I propose that we can retain such an approach, but to do so we must admit *all three* interpretations of "will" into the mix. Given that each interpretation implicates different exempting conditions, though, our responses to them are going to be best construed as tracking different *conceptions* of responsibility, what I have elsewhere labeled attributability, answerability, and accountability.[27] What I will propose, in other words, is a *qualities* of will theory of responsibility according to which our ambivalent responses to marginal agents reflect our commitment to the thought that these agents are responsible in some ways but not others. I will argue for this proposal by taking much more seriously than has been done before what the whole range of our responsibility responses consists in, what precisely these responses target, and what capacities they presuppose. If we can account for our ambivalent responses by going this route, then Watson's worry may finally be laid to rest.

IV. The Wide Array of Responsibility Responses

So what are our responsibility responses? These are typically put in Strawsonian terms as our *reactive attitudes,* what Strawson took to be the affective attitudes we have to the attitudes of others. Most theorists have focused primarily (sometimes exclusively) on the negative reactive attitudes of resentment, indignation, and guilt. But Strawson mentions additional negative reactive attitudes that are often ignored but that nevertheless also implicate responsibility, namely, hurt feelings,[28] and shame.[29] Beyond Strawson, though, there are negative reactive attitudes that seem to fall squarely within our responsibility practices, including disdain, contempt, disapproval, regret, condemnation, embarrassment, and disappointment. And it is equally important, albeit less often recognized as such, that we include the wide range of *positive* responsibility responses. The primary

[26] For discussion of some of the previous cases, see my "Moral Address, Moral Responsibility, and the Boundaries of the Moral Community," *Ethics* 118 (2007): 70–108; as well as my "Responsibility and Disability," *Metaphilosophy* 40 (2009): 438–61.

[27] David Shoemaker, "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics* 121 (2011): 602–32.

[28] Strawson, 75.

[29] Ibid., 85.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

one that Strawson mentions is gratitude. But beyond gratitude we must include pride, admiration, approval, and warm feelings.

These emotional attitudes have as a constitutive component—the component to which we often point to identify and differentiate them—various motivational and behavioral manifestations.[30] Admiration, for example, typically includes a motivation to emulate the admired person or improve oneself.[31] Shame typically includes a motivation to hide from the gaze of others.[32] Guilt typically includes a motivation to repair relations.[33] And there are many other behavioral manifestations of the large variety of our reactive emotions. On the negative side, there may be shunning, criticizing, haranguing, withdrawing friendly greetings, modifying dispositions for trust and cooperation, quietly ending friendships, and more.[34] And on the positive side, there may be praising, rewarding, increased friendliness, increased trust and cooperation, and more.

There are thus many responsibility responses we might have to others. What I have argued thus far is that, given our ambivalent responses to different marginal cases, these responses cannot be a unified set targeted to only one of the types of will on the table. Indeed, they may seem more like a fractured grab bag. Nevertheless, what I will now argue is that our responses are quite plausibly grouped into three distinct categories: some target character, some target judgment, and some target regard.

V. The Umbrella Concept of Responsibility and Our Agential Cares

I have discussed several types of emotional responses that I have put under the rubric of responsibility responses. But what unifies disdain and regret and resentment and gratitude and admiration and so forth as *responsibility* responses? In general, we have responsibility responses only to characteristics of *persons* (where these are not restricted to humans). But not just any characteristics of persons generate responsibility responses.

[30] I am grateful to Michael McKenna for impressing upon me the advantages of putting the point in these terms. See also his *Conversation and Responsibility,* 69–70. I have also benefited here from conversations with Justin D'Arms, Daniel Jacobson, and Shaun Nichols.

[31] See, e.g., Mary Helen Immordino-Yang and Lesley Sylvan, "Admiration for Virtue: Neuroscientific Perspectives on a Motivating Emotion," *Contemporary Educational Psychology* 35 (2010): 110–15; and Sara B. Algoe and Jonathan Haidt, "Witnessing Excellence in Action: The 'Other-Praising' Emotions of Elevation, Gratitude, and Admiration," *Journal of Positive Psychology* 4 (2009): 105–27.

[32] See, e.g., Bernard Williams, *Shame and Necessity* (Berkeley, CA: University of California Press, 1993); and Gabriele Taylor, *Pride, Shame, and Guilt: Emotions of Self-Assessment* (Oxford: Oxford University Press, 1985).

[33] See, e.g., Brian Lickel et al., "Vicarious Shame and Guilt," *Group Processes & Intergroup Relations* 8 (2005): 145–57; Mia Silfver et al., "The Relation Between Value Priorities and Proneness to Guilt, Shame, and Empathy," *Motivation and Emotion* 32 (2008): 69–80; and Brian Parkinson and Sarah Illingworth, "Guilt in Response to Blame from Others," *Cognition & Emotion* 23 (2009): 1589–1614.

[34] See Scanlon, *Moral Dimensions,* ch. 4.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

For instance we do not have such responses to people's beauty or height (other things equal). And being a person is not sufficient to generate responsibility responses. So what do these responses track? They all track expressions of *practical agency,* the actions or attitudes of entities with volitional structures who reflect on what to do and what attitudes to have. Our responsibility responses target not just persons but agents.

Now we care about our fellow agents, of course, but we also care about the *agency* of our fellow agents. In particular, we care about three specific agential features: the *kinds* of agents they are, the *decisions* they make, and the *regard* they have for us. That is to say, we care about their motivational character, how they govern themselves, and how they view us. These are the three different targets of our PQW theorists, of course. But we clearly care about *all three.* And they are quite distinct targets, each with different satisfaction conditions.

We can best see the differences between these targets in terms of our distinct affective responses to them, derived directly from our caring about them. Caring, among other things, is emotionally dispositive: to care about X is to be disposed to respond emotionally to, and in sync with, the up-and-down fortunes of X.[35] If I care about you, I tend to be pleased when I perceive your fortunes to be good and upset when I perceive them to be bad. So when I care about a feature of your agency, I am disposed to respond emotionally to, and in sync with, its good or bad expressions. Here, then, is the thought I will try to develop in what follows. There are three distinct subsets of responsibility responses typically triggered by different qualities of will, and for each subset there are paradigm emotional pairings (positive and negative). The paradigm responses to qualities of character are admiration and disdain (although there are many others, including contempt and shame). Paradigm emotional responses to qualities of judgment are regret and pride (although again, there are many others, including disappointment, disapproval, and embarrassment). Paradigm emotional responses to qualities of regard are anger and gratitude.[36] Each, further, has typical behavioral manifestations. These distinct syndromes of response to different qualities of will

[35] See my "Caring, Identification, and Agency," *Ethics* 114 (2003): 91. Cf., Harry Frankfurt, *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999), e.g., 159–62; and Agnieszka Jaworska, "Caring and Internality," *Philosophy and Phenomenological Research* 74 (2007): 529–68.

[36] The astute reader will have noticed that regret and pride are first-personal, whereas the paradigm emotional pairings attached to character and regard are third-personal. This is primarily because the third-person responses we have to the quality of other people's judgments typically stem from our caring about our own judgments: we care about how others decide primarily as a way to improve how we decide. (It matters to me what Jesus would do primarily insofar as it helps me figure out what *I* should do.) Given the (typically) derivative nature of our caring about others' judgments, then, I think it best to discuss our paradigm responses in this zone as being targeted to our own judgmental qualities. For those bothered by the asymmetry to the other qualities of will, however, I say more to motivate and defend it in *Responsibility from the Margins,* unpublished book manuscript.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

are the basis of the three conceptions of responsibility: attributability, answerability, and accountability. In the remainder of the paper, I will sketch each conception in light of our responses, and I will attempt to show just how each conception differs from the others by paying close attention to those responses.

VI. Attributability and Quality of Character

For me to be responsible for Φ in any sense, Φ must be *attributable* to me. But some actions or attitudes, it seems, may be attributable to me in one sense but not another. For example, those with Obsessive Compulsive Disorder (OCD) or kleptomania perform actions determined by "their" intentions, in the sense that they are determined by motivating attitudes internal to their psychic domain (and not someone else's); but in another sense these intentions are not "theirs" in virtue of not being governable by any of the features we think are essential to them as *agents.* One way of putting this that some theorists have found amenable is that these attitudes are not governed by the agent's *deep self.*[37] For purposes of responsibility, then, an action or attitude is attributable to one just in case it depends on, or expresses, one's deep self.

Lots of ink has been spilt about the nature of the "deep self." The idea is to figure out just what psychic elements are most deeply representative of the agent. The controversy has been over whether the deep self is found in some noncognitive, Humean strand (e.g., cares) or some cognitive, Platonic strand (e.g., evaluative judgments).[38] These strands are also called, respectively, "authenticity" and "authority" conceptions of the deep self.[39] But I see no reason to choose sides. Sometimes, it seems, one's cares determine one's actions independently of one's evaluative judgments, and sometimes the reverse is the case. An example of the former is Harry Frankfurt's case of volitional necessity, occurring when someone cares about something in such a way that he is unable to bring himself to betray the care, and he would not *want* to betray it, regardless of his judgments of the intrinsic value of doing otherwise.[40] And examples of the latter are rampant: I may continue writing, despite caring a lot about playing golf instead, because I judge it would be better to do so.

[37] See, e.g., Susan Wolf, "Sanity and the Metaphysics of Responsibility," in Watson, *Free Will,* 372–87. For other theorists, this way of putting it will be more irritating than helpful, but I maintain it for ease of exposition, and they can substitute it with talk of agential and non-agential attitudinal causes without any real loss.

[38] For the classic statement of the former, see Harry Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), 11–25, 58–68, and 159–76. For the classic statement of the latter, see Watson, *Agency and Answerability,* 13–32.

[39] See Kasper Lippert-Rasmussen, "Identification and Responsibility," *Ethical Theory and Moral Practice* 6 (2003): 349–76.

[40] See, e.g., Frankfurt, *The Importance of What We Care About,* 85–88.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

But identifying what the deep self consists in is less pressing at this point than determining whether attributability is indeed sufficient for *responsibility.* For our purposes, that amounts to determining whether attributability is sufficient to generate some among our set of responsibility responses. And it is, regardless of whether the deep self is located in one's cares or one's evaluative judgments. But the story is a bit complicated. Whether some attitude is attributable to one's deep self is a structural fact. But why should the mere fact that some of one's attitudes (e.g., intentions) are dependent on or express some of one's other attitudes (cares or evaluations) generate responsibility responses? It is because when they bear the right relation, something is revealed of *who the agent is qua agent.*

Consider what would happen were you to run across someone who insulted everyone he met in the following brutally succinct way: "Stupid!" "Ugly!" "Dumbass!" "Fatso!" Suppose further that you are inclined to give people the benefit of the doubt, so you initially think that he must have Tourette syndrome. When you ask me about this person (whom I know), I respond that no, he is not Tourettic. I suspect you will immediately feel (at the least) disdain. Why? It is because you suddenly realized that his actions actually expressed *his* attitudes, attitudes that depended on what you now take to be his character, and not on dysfunctional brain circuitry, and so you have actually come to know something about who he really is: he is a jerk, and disdain is surely appropriate for jerks.

Or consider discovering someone you had thought of as aloof and uncaring helping some poor local kids build a basketball court, something he has been doing for quite a while, as it turns out. The most natural emotional response would be admiration. This emotion has a certain feeling—typically described as a sensation of uplift—with a motivational impulse to emulate the admired person and improve oneself.[41] What you are admiring is excellence, specifically the excellence of his *character.* You are now responding to who you take him to really be *qua* agent. But then suppose you find out that he has been hypnotized to do these things, or that he has a brain tumor that causes such behavior. These actions and attitudes are no longer attributable to him, and admiration thus no longer seems appropriate, precisely because nothing about *him* is any longer being revealed.[42]

The most familiar plea to exempt agents in virtue of their lack of attributability-responsibility is "She's just not herself." Perhaps the most

[41] See, e.g., Algoe and Haidt.

[42] It may, nevertheless, linger, being "stably recalcitrant" in the words of D'Arms and Jacobson. See their "The Significance of Recalcitrant Emotion (or, Anti-quasijudgmentalism)," *Royal Institute of Philosophy Supplement* 52 (2003): 127–45. This is one of the reasons I think of admiration as a sentiment, an irruptive, pan-cultural emotion with a syndrome of associated thought, feeling, and motivation, one that evaluates objects independently of judgment (see Justin D'Arms, "Value and the Regulation of the Sentiments," *Philosophical Studies* [2012], DOI: 10.1007/s11098-012-0071-9). This is a controversial position, though, and one that I do not defend here. See *Responsibility from the Margins* for detailed discussion and defense.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

heartbreaking example of this paradoxical plea comes in cases of clinical depression. The thought is that the agent we see before us is either not the sort of agent to whom certain attitudes—for instance, apathy—ought to be attributed or not the *same* agent to whom we are accustomed to attributing these attitudes. These are agents who lack the relevant quality of character to generate our *aretaic* emotional responses. There are then at least three capacities necessary for attributability-responsibility: the agent must have (a) a deep self (the capacity to care or evaluate), (b) the capacity to have the relevant attitudes depend on or express her deep self, and (c) a certain sort of persisting agential identity.[43]

This last condition is somewhat obscure, but it is intended to capture the fact that, from the third-person perspective, the emotional responses targeted to agential character are best informed by behavioral *patterns*, so the agent to whom *aretaic* responses apply must persist in some sense for these patterns to reveal traits. The most reliable way to determine what someone's character consists in is to see how that agent behaves in similar circumstances on a number of occasions. Of course, someone who gives a lot of money away at once, or someone who tortures a child may give us all the evidence we need about his character in that one-shot action. But there are clearly some character traits revealed *only* by patterns of behavior. Think of someone who is a "ditherer," or "flighty," or "undependable," or "stingy." The responses we have to them are not to any of their specific actions or attitudes; instead, they are responses to patterns of behavior or attitudinal expressions we perceive to be attributable to them. And this is also why I think it is a mistake to pick either cares *or* evaluations as the sole source of the deep self, for we also have responsibility responses to people in virtue of their being motivated mostly by their cares or being motivated mostly by their evaluative judgments (think, for example, of those who are "impulsive" or "blindly loyal," as opposed to those who are "calculating" or "deliberative"). Either sort of pattern is revelatory of the agent's deep self.

Attributability is thus, I think, sufficient for responsibility, or more specifically, one *conception* of responsibility. Our responsibility responses in this zone—the paradigm ones being admiration and disdain—target agential character, which is revealed when a certain structural relation obtains between actions or attitudes and the agent's deep self. When that relation obtains, one's character is thought to be expressed in those actions or attitudes. Our *aretaic* emotional responses are to the quality of that character, though, not to the structure that reveals it.

VII. Answerability and Quality of Judgment

Attributability answers yes to the question, "Did *you* do it?" where the probe is for whether the intention producing the action itself had the right

[43] I take no stand here on whether these are jointly sufficient.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

connection to one's agency. Answerability, the second conception of responsibility, consists in one's being in principle able to answer the familiar question, "*Why* did you do it?" In asking it, we are trying to find out from the agent what he took his justifying reasons to be in performing some action or having some attitude. In particular, we want to know why he apparently judged some reasons to be more worthy than others. To be answerable for Φ (an action or attitude) thus presupposes attributability *of one kind*, namely, the kind wherein Φ is governed by one's evaluative judgments—that Φ is, in Scanlon's words "judgment-sensitive."[44]

Now it has seemed to some PQJ theorists as a result that answerability *just is* attributability.[45] This is not so, however, for a few reasons. First, as I have just indicated, one is answerable for one's attitudes only if they are governed by the *evaluative* strand of attributability. But one is not answerable for attitudes flowing ultimately from one's caring strand, where these cares are independent of one's evaluative judgments. In illustration, think of actions motivated by unconditional parental love. While I, a parent, may be able to cite the *explanation* for why I continue to bail out my good-for-nothing son—I love him—that is no *justification* (I judge him to be worthless, after all). My "answer" is not of the right sort to make me answerable.

Second, attributability is just about whether some action or attitude really reflects your general motivational character, so any single action or attitude offers only limited evidence for that ultimate target. But answerability applies exclusively to individual actions or attitudes themselves. We want to know why you took yourself to be justified in doing or feeling this specific thing, *regardless* of whether it reflected some character trait.[46] That there is a deep volitional structure in place is less important on this conception of responsibility than what the ultimate source of the current structure is, namely, whether it is a judgment about reasons pertaining to a specific action or attitude.

This point is borne out by the set of responses relevant in this realm, which are responses only to actual (or presumed) *answers* about one's judgments. (So answerability as such, just like attributability as such, does not yet render any particular responses warranted). The paradigm paired emotional responses here are first-personal: regret and pride. There are also third-person analogues of these emotions: I may be disapproving, disappointed, or cringe with embarrassment on your behalf when you have acted on bad reasons, and I may be proud or even awestruck when you act on excellent reasons. Behavioral manifestations include attempted

[44] See, e.g., Scanlon, *What We Owe to Each Other,* 18–22.

[45] In addition to Scanlon in the previously cited works, Angela Smith explicitly advocates such a view in, e.g., "Attributability, Answerability, and Accountability: In Defense of a Unified Account," *Ethics* 122 (2012): 575–89.

[46] For those who act from the heart, who are impulsive or guided by unconditional love, actions flowing from their evaluative judgments may actually be out of character.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

policy changes, in the first-person case, and criticisms, compliments, or altered dispositions of interaction, in the third-person case. The target of all these responses is the quality of your judgment about reasons, as revealed by the decisions made and intentions formed. What renders your regret appropriate, for example, is that you made a terrible decision, mistakenly or foolishly judging that the wrong reasons were more worth adhering to than the right ones. This is a failure of self-governance.

I have already noted why answerability is distinct from attributability: some attitudes may be deeply attributable to one without one's being answerable for them, as in the case of attitudes dependent on or expressive of one's non-judgment-sensitive cares. But what really distinguishes the two is the focus in answerability on the character of the specific judgment made, not on the character of the agent making it. Our ambivalent responses sometimes reflect just that distinction: I may continue to admire *you* while being disappointed in how you have decided. Conversely, I may have nothing but disdain for you while still offering grudging compliments for some decision.

But why haven't I included the paradigm "blaming" attitudes? Isn't resentment, say, an appropriate response to poor quality of judgment, in particular to poor *moral* judgment? Not in and of itself. We exercise judgment across all sorts of practical domains, and there is nothing about the quality of our judgment alone that changes in the different domains. So poor quality of judgment in the athletic realm is, in and of itself, exactly like poor quality of judgment in the aesthetic, prudential, *and moral* realms. But if resentment is inappropriate in the non-moral realms, and there is nothing intrinsically different about quality of judgment across them all, then resentment would also seem to be inappropriate for poor judgment on its own in the moral realm.

In illustration, consider the following hypothetical:

> *American Moral Idol (AMI):* Contestants from around the country vie to participate in this reality show, in which the winner will be given a million dollars and a certificate as a new moral paragon. Each week on the competition, contestants are dropped into morally charged scenarios in which they must decide how to act. They are then critiqued by three famous judges, the most curmudgeonly (but most trusted) of whom is Simone Powell. One week, two contestants, Clay and Carrie, are each placed on a street in which they find a pair of crutches. They can sell the crutches at a pawn shop across the street and keep the money. But in between them and the store is a young woman who has a cast on her broken leg and is hobbling about. Both contestants deliberate and then buzz in to announce their decisions. Carrie decides to offer her crutches to the girl with the broken leg. Clay decides to cross the street in order to get his pawn shop refund. After their decisions have been announced, the judges make their

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

> standard answerability demand ("Why did you make *that* decision?"). Carrie's response is that the woman needed the crutches more than she, Carrie, needed the money. Clay's response is "Duh, free money!" Simone Powell sharply criticizes Clay: "What you decided to do completely ignores this poor girl's need for those crutches, favoring instead the piddling amount of money you'd get from the pawn shop. What a terrible decision!" Only Carrie makes it to the next round.

Simone has no personal stake in the matter, and she is genuinely interested in helping America find the next true moral idol. Her moral criticism of Clay is right on the money. But there is nothing distinctive about it *qua* criticism, i.e., it is no different from criticism of poor judgment in any other normative domain.[47] It also does not consist in, or render appropriate, Powell's resentment or moral indignation. At most what it renders appropriate is disappointment or disapproval on Simone's part, and regret or embarrassment on Clay's, just like poor judgment of other kinds.

Scanlon, however, disagrees:

> If an action is blameworthy, then the agent has either failed to take account of or knowingly acted contrary to a reason that should, according to any principles that no one could reasonably reject, have counted against his action. So, in addition to whatever loss this action may have caused, the agent's mode of self-governance has ignored or flouted requirements flowing from another person's standing as someone to whom justification is owed. This is what makes the action wrong rather than merely harmful, and *it is what makes it appropriate for the person who was wronged to feel resentment rather than merely anger and dismay.* . . . What is special about violations of the morality of right and wrong is that the reasons one has failed to respond to are grounded not just in some value that others also recognize but in *their own* value as rational creatures. These violations therefore have particular importance for one's relations with them.[48]

I do not find this claim persuasive. If we owe justification to others in virtue of their standing and value merely as *rational* creatures, solely as creatures capable of recognizing, assessing, and being moved by reasons, then why should we think that the *kind* of reasons being flouted render appropriate unique *moral blaming*, the resenting response? In other words, why don't you fail to recognize my value as a rational creature every time you

[47] Imagine, for instance, an American Chess Idol, or the actual American (singing) Idol, in which this sort of criticism regularly takes place in just this way.

[48] Scanlon, *What We Owe to Each Other,* 271–72; first emphasis mine; second emphasis in original.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

perform an action for poor reasons, period? Part of what matters about our relations with one another, after all, is that you and I govern ourselves in ways that make sense, in ways that are rational; otherwise, we cannot coordinate our activities and even *have* fruitful interpersonal relationships. Irrationality is the worst form of unpredictability and it can even render people unintelligible to us. But presumed unintelligibility in nonmoral behavior generates just as pressing a demand for justification as presumed (immoral) behavior; that is, you owe me a justification just as much for your seemingly crazy decisions as your seemingly nasty ones.

As a way of responding to a worry along these lines, Scanlon asserts that "moral criticism is a special case of . . . more general rational criticism," insofar as "[m]oral criticism claims that an agent has governed him- or herself in a manner that cannot be justified in the way *morality* requires, and it supports demands for acknowledgment of this fact, and for apology, or for justification or explanation."[49] In other words, goes the thought, morality requires regard, and disregard renders resentment and such appropriate. But by Scanlon's own admission, what falls under the rubric of morality may be far more than the regard-based demands of what we owe to each other.[50] So sexual impurity, or dishonorable actions, or failures to develop one's talents may all count as regrettable and *morally* criticizable without calling for anything different from nonmoral, rational criticism (and so not blame). And even when we restrict our focus to the subset of regard-based morality, mistaken moral judgments alone do not yet render the blaming emotions appropriate, as illustrated by Clay in *AMI:* he makes a mistake in failing to properly regard the interests of the hobbling young woman, but to this point he has only made a criticizable *judgment* about the worth of the various reasons, and this alone is insufficient (at least in this case) to gin up blame. What renders such emotions appropriate in addition is the subject of the next section.

The criticisms or compliments manifesting our emotional responses to quality of judgment exclusively target the way agents govern their relation to *reasons;* and so despite what Scanlon and other PQJ theorists want to say, answerability, while capturing a lot of our responsibility responses, does not capture them all. It fails on the one hand to capture our responses to character-based expressions of nonrational cares; and it fails on the other hand to capture an essential set of responses targeting how agents relate, not to reasons, but to *other agents,* our final topic.

VIII. Accountability and Quality of Regard

The typical third-person emotional response targeted to poor quality of character—disdain—grades it as falling short of some *aretaic* ideal.

[49] Ibid., 272; emphasis mine.
[50] Ibid., 173.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

The typical third-person response targeting poor quality of judgment—disappointment—grades decisions made as falling short of some judgmental norm. Neither of these responses requires confronting their target, however. We make quiet assessments of responsibility of both kinds all the time, and what we do more often than not is then simply adjust our lives accordingly with respect to the lives of our targets. Judgments of their responsibility, even blameworthiness, do not commit us to expressing anything *to them,* even if various emotional responses to who they are or how they judged are appropriate.

But there is one emotional response to poor quality of will that we have not seen applied yet, and it is the one that most theorists have taken to be the paradigmatic responsibility response: *anger.* Indeed, this is the response that theorists typically think of as the basis of blaming sanctions, consisting in the harsh treatment that lends the issue of moral responsibility its own air of moral urgency. Now most theorists these days do not talk about anger. Instead, they talk about Strawsonian resentment or indignation. These are just variations on anger, though, restricted, for instance, by thoughts about the reactor's relation to the injuring agent. My own view is that these restricted labels obscure some important points, but I will not go into those here. Rather, I want to focus on why anger's role in moral responsibility has been misunderstood, and it is my aim in this section to explain what I think its role actually is, as well as why it motivates my talk of yet another conception of responsibility—accountability—which implicates a distinct set of capacities.

We are interested, then, in the anger directed at other agents; call it *agential anger.* Here, as in many other matters, Aristotle was on the money:

> Anger may be defined as an impulse, accompanied by pain, to a conspicuous revenge for a conspicuous slight directed without justification towards what concerns oneself or towards what concerns one's friends.[51]

There are many elements worth commenting on here, but I will focus on just two. First, agential anger responds to slights. This is a response, then, to poor quality of *regard* of some kind. But what kind is it, precisely, and why wouldn't our angry responses to slights fall under the same category as responses to poor quality of judgment? Second, anger is defined as an impulse to a conspicuous revenge. What is the role of conspicuousness here, though, and how should we think of its relation to revenge?

Start with the first issue. Aristotle often talks of slights as insolence. This seems too restrictive, though. When you talk loudly to your neigh-

[51] Aristotle, *Rhetoric* 1378a; from a hyptertext resource compiled by Lee Honeycutt from the translation by W. Rhys Roberts. URL: http://rhetoric.eserver.org/aristotle/.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

bor throughout a flight, you are, it would seem, slighting all of the other fellow passengers who are trying to work or sleep. But you are not being insolent. Insolence is a matter of active disregard, whereas slights also include "mere" failures of consideration, a matter of *insufficient* regard.

Perhaps, then, we should go with the *au courant* understanding of the object of Strawsonian resentment, namely, *wrongings.*[52] If this is correct, then slights—insufficient regard—would seem to consist in a failure to take sufficiently seriously second-personal *reasons,* agent-relative reasons expressing legitimate claims we make on one another.[53] But then wrongings would consist in bad decisions about the worth of those second-personal reasons; that is, they would fall under the rubric of an exercise of poor quality of judgment and so could be gathered under the answerability tent.

I now believe that this is an incorrect account of the appropriate object of anger.[54] To see why, let me first note that slights range more widely than mere wrongings. There are many examples here. Take suberogatory actions: I have no claim on you not to mow your lawn at 8 A.M. on a Sunday morning, but if I have told you repeatedly how much I dislike it, it would sure seem as if you slight me when you continue to do so.[55] Or suppose that I have brought only one pencil to an exam whereas you have brought twenty. When my lone pencil breaks, I may look to you with a pleading face. I have no claim on any of your pencils. Nevertheless, suppose you look back at me with a smug grin and then return to your exam, failing to give me a single one. This too seems to be a slight, one that appropriately generates anger on my part.[56] And the same goes for someone who refuses to give his dying brother one of his kidneys to save his life: the dying brother has no such claim, but he nevertheless would be appropriately angry at the slight.[57] All of these are violations of "mere" *expectations.* They are not claims, the demands expressing second-person reasons, but they are also not just hopes or wishes. They articulate behavior we expect from our fellows: basic consideration, painless generosity, or brotherly sacrifice. Those who fail to meet these expectations may slight us just as those who wrong us may.

[52] There are many who hold such a view. The most prominent include Wallace; and Stephen Darwall, in *The Second-Person Standpoint* (Cambridge, MA: Harvard University Press, 2006). I too have explicitly advocated this view in several papers, including "Attributability, Answerability, and Accountability," as well as "Moral Address, Moral Responsibility, and the Boundaries of the Moral Community." As will now become obvious, I no longer hold this view.

[53] See, e.g., Darwall.

[54] In part because of McKenna, from chap. 8 of his *Conversation and Responsibility.*

[55] Drawn from a case Julia Driver explores in "The Suberogatory," *Australasian Journal of Philosophy* 70 (1992): 288.

[56] Drawn from a case Judith Jarvis Thomson explores in *The Realm of Rights* (Cambridge, MA: Harvard University Press, 1990), 117.

[57] Driver, 287; McKenna, *Conversation and Responsibility,* 184.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

But even if we include expectations alongside wrongings as slights, why aren't we still just pointing out that those who slight us exercise poor quality of *judgment:* they make bad decisions about the worth of various reasons, either from demands *or* expectations? The answer is that slights may occur even in the absence of wrongings or expectational letdowns. Suppose that I help you solely because it is my Christian duty, or because I am trying to impress someone, or because I am in the process of becoming a Nietzschean Uberman. In each of these instances, your role as the *source* of the relevant expectation has been entirely disregarded. To help you only because God demanded I do so, for example, leaves you—the one who could get hurt, after all—looking like chopped liver.

Now the Scanlonian might reply that this is still solely a matter of poor quality of judgment, given that our expectations incorporate certain motivating reasons. For example, I have an expectation that you will help me *because I expect it*, and not for other reasons. While I think this understanding of expectations would be far too restrictive, nevertheless it is possible to imagine cases in which even such expectations could be met with slight. One example comes from *how* I might meet them. Suppose that I meet your demands or expectations precisely because they are yours, but I do so ultimately because it amuses me. You are such a lowly person, I may think, and I find it hilarious to see you running around giving your betters orders. I am insufferably condescending, and so subject to your wrath.

But perhaps my condescension is just an attitude that violates a *different* demand or expectation of yours, namely that I not be condescending, and that is why my method of adhering to your *other* demands or expectations is anger-appropriate? But this would just amount to a demand that I take your demands seriously, and as such it could not get off the ground. If I do not already take your demands seriously, then your meta-demand gives me no reason to do so.

The problem is that I do not take *you* seriously. Until I do so, I can at most be construed as merely following your orders. And what is required for me to take you seriously? Your normative perspective on the world must *matter* to me. The specific sorts of attitudes and treatments of you that you consider good or poor must count for me too—they must carry normative weight—and they must count in sync with the way they do for you.[58] This mattering may be a function of my judging you to be valuable, but it could just be a function of my caring about you.

I slight you, then, when your normative perspective bears insufficient weight for me. This may happen in one of two ways. In the first, I may mistakenly judge that the reasons against treating you in a certain way,

[58] Psychopaths, after all, may take seriously my perspective about what pains me, but only insofar as it provides them with information on how best to pain me. I have put this point elsewhere ("Moral Address, Moral Responsibility, and the Boundaries of the Moral Community") as motivating the requirement that we *share the cares* of the other.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

reasons provided by your value as a fellow rational creature, are insufficiently weighty. This is the kind of disregard Scanlonians take to be already included under the rubric of moral answerability. I am insisting instead that it is important to keep regard and judgment prized apart, at least conceptually, for it is only the poor quality of regard aspect that renders *anger*—holding accountable—appropriate, and not the poor judgment aspect in and of itself. But at any rate, this is only one type of disregard. The other type occurs when I simply fail to *care* about your perspective, where this translates into my not being *emotionally* affected by your fortunes (in sync with you). I slight a loved one when I am not naturally upset by what has upset her, but also when I am not happy at what has made her happy. I may judge her to be valuable, of course. But judgments of value have no necessary ties to emotional susceptibility, not even (it would seem) for the ideally rational: just as I may judge that avant-garde jazz or the compositions of Schoenberg are valuable—and merit respect—without being touched by them at all emotionally, so too may I judge that other people are valuable—and that their interests merit respect thereby—without being at all emotionally responsive to their plights. Yet an emotional component—the expression of non-judgment-sensitive caring—is, it seems, essential to many forms of interpersonal regard. Its lack may thus generate slights even if the slighter makes excellent judgments about the worth of the reasons flowing from the slighted's value as a fellow rational creature.

A slight, then, is at bottom a failure of empathy.[59] Anger responds directly to slights to oneself or others,[60] and these may reflect either the slighter's evaluative stance *or* his nexus of cares. This is why slights range more widely than wrongings and expectational letdowns, for they may also consist in the absence of a kind of emotional susceptibility implicating attitudes beyond the judgment-sensitive.[61]

Turn now to the second issue. Aristotle calls the angry response to slights an impulse to "conspicuous revenge." What are we to make of this? Here I want to argue that it is the *conspicuousness* that matters more than the revenge. The basic thought is this: *to be communicated* is anger's constitutive motivational component. One simply does not count as being angry at someone without having a motivational impulse to communicate that anger to the slighting party. This feature may be quashed, of course; if Mike Tyson is the one who slighted me, I will surely do my best to keep my face a smiling, polite mask. But the point is that there still is

[59] Compare and contrast the discussion in Darwall, 43–48. He focuses solely on the recognitional, but not the emotional, elements of empathy.

[60] I lack space to discuss how this account of anger for slights to oneself might be generalized, very naturalistically, to anger for slights to others, but I attempt to do so in *Responsibility from the Margins*. The story appeals to what I think is a natural empathic extension.

[61] Again, I expound on these ideas in *Responsibility from the Margins*.

an impulse to communicate my anger that has to be *repressed*. And bottled up anger can be deeply corrosive, eating away at one's insides, creating obsessions, nightmares, ulcers, and ground-down teeth. It is also enormously difficult to repress for long. Indeed, release—that is, expression—of that anger often yields relief, the removal of a burden, as the anger has finally been set loose to fulfill its function.

Getting back at the slighter, achieving revenge or retribution is the most powerful and obvious form of communicating one's anger. But there are many other ways to communicate it: shunning, badmouthing, yelling, giving the silent treatment, and more. Even people who are motivated to the most extreme form of revenge—killing—feel the need to let their victims know why: "This is for what you did to me (or my dead wife or sister, and so on)!" This is why I think the conspicuousness of the anger response is more fundamental than the revenge component, because the latter *serves* the former. Revenge without communication of the anger motivating it is deeply unsatisfying and starts to feel less like anger and more like merely a cold meting out of justice.

Once we start talking about revenge as part of anger, however, moral flags go up. Indeed, it is typically assumed that because the aim of the so-called retributive reactive attitudes (such as resentment and indignation) is to sanction, they must meet fairness conditions, such as the ability of the agent to avoid the sanction.[62] The premise is false, though, at least with respect to the blanketing sentiment: anger's aim is not necessarily to sanction (nor do I think sanction is the aim of resentment or indignation, to the extent that they are any different). While revenge is often on the mind of victims of slights, it is not a necessary component of their anger, as they may merely be motivated to shun or to yell at the slighter. *Communication* of anger is built in, but such communication does not aim at sanction. Instead, it simply aims at *making the slighter aware of his slight,* to get him to fully appreciate his failure to take sufficiently seriously the slighted agent's normative perspective and so to feel what it was like to have been disregarded in that way (and I will say more about what this consists in below). This communication may hurt the slighter, of course, just as many other sorts of communications may hurt (e.g., "I find your novel trivial," "The dinner you have prepared is pretty bland," "You call that a jump shot?!"), but such hurt is merely a contingent side effect of anger's fundamental motivational function, which is just to get the slighter to robustly acknowledge the slighted's perspective both on what was done and how he felt about it.[63]

This point is brought out by considering the rarely discussed positive cases in this arena. There is no term for the opposite of "slighted," but I

[62] See, e.g., Wallace, Watson (*Agency and Answerability,* 279–80), and Shoemaker ("Attributability, Answerability, and Accountability"). Again, I have changed my view.

[63] See McKenna, *Conversation and Responsibility,* 141–46, for a discussion of the harms of holding accountable as mere side effects.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

think we clearly recognize its occurrence. To slight someone is to take him (his normative perspective, his interests) insufficiently seriously. For the opposite I will coin a phrase: to *height* someone is to take him (his normative perspective, his interests) *very* seriously. And as being slighted renders fitting anger, being heighted renders fitting *gratitude.*[64]

Gratitude is not fitting for the mere meeting of demands or obligations (unless, perhaps, the difficulty of doing so is significant). Instead, it operates in the world of expectations, namely, when one's expectations are treated like demands, or when one's mere hopes or wishes are treated like things one actually has a right to expect. The heighter takes one's normative perspective and interests quite seriously, and as with slighting, the heighter may height one either by (a) judging that the reasons provided by one's expectations or hopes count more *heavily* than even the heighted agent thinks they do, or (b) emotionally identifying with the heighted to an unexpected (but welcome) degree. Analogous with slighting, then, the first sort of heighting has its source in the heighter's evaluative stance, whereas the second may have its source in the heighter's non-judgment-sensitive nexus of cares. Examples of (a) are fairly obvious. Suppose after taking up your perspective I come to fully appreciate just how much you would love to meet Tom Waits, so at great expense I arrange for it to happen. I have judged that the reasons in favor of granting your (mere) wish count for more than even you think they do (it was, after all, just a *wish,* nothing you expected ever to be granted). Your gratitude at my heighting is quite fitting here. But suppose you have been waging a lonely fight for justice in some small corner of the world, and after a serious setback to your cause, I, just a quiet observer of your fight, am emotionally wrought alongside you. I have heighted you, and your gratitude is fitting, insofar as you can see that your loss hurts me *as if it were my own,* in an unexpected but welcome way. I have come to care about what happens to your cause *on your behalf,* it turns out, more than you had any reason or right to expect.

Notice, next, that gratitude, just like anger, has a *to be communicated* component built-in. I do not count as grateful unless I also have the motivational impulse to communicate that gratitude, to make it *conspicuous,* either through effusive thanks, or "good-mouthing," or doing favors in return, or sending a card, and so on. And while there may be reasons not to communicate one's gratitude (perhaps it will just go to the heighter's head), the impulse to do so regardless must be repressed, which takes some effort. Note also that gratitude, as the opposite of anger, does not necessarily have as its motivational aim to benefit or help the heighter. Again, the aim is fundamentally communicative, to get the heighter to robustly acknowledge what he has done and how he has made the heighted feel from the heighted's perspective. But if the aim is to communicate, not to benefit, fairness considerations seem irrelevant here as well.

[64] Scanlon talks about gratitude as the opposite of resentment in *Moral Dimensions,* 151–52.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

So what is the content of the communication, then, and how does it implicate what we would think of as the capacities for responsibility? I will focus once again on anger, the content of which, I think, is two-fold. First, it is essentially *emotional,* expressing how the angry agent feels about what was done. Second, it is the expression of a demand, but not, either fundamentally or even necessarily, the basic demand for goodwill (as I have previously thought, following Watson).[65] Rather, it is the expression of a demand *to take up the perspective of the angry party and to feel what it was like to be her when slighted.* Call this the demand for identifying empathy: it requests the slighter to identify with the slighted emotionally. What is expected from this exercise is that the slighter will, first, feel what it was like to be the slighted—to experience the other's hurt and subsequent anger—and then, second, return to his own perspective and have the anger he just felt from the other's perspective transmogrify into *guilt.*

What is fundamentally demanded, then, is robust emotional acknowledgment on the part of the slighter: the slighted is making a dramatic, emotional demand of him that he come to recognize how he disregarded the slighted and that he share with the slighted the emotional consequences of that disregard. The aim is thus not primarily to engage in moral dialogue, nor to produce reintegration of the slighting agent into the moral community.[66] It is just to make the slighter dramatically aware of what he has done. It is thus essentially backward-looking.[67]

Once we understand quality of regard in this way and see it as the target of its own class of responsibility responses (variations on anger and gratitude), we can see why there is a third, distinct, conception of responsibility: *accountability.* Being an accountable agent requires the capacity to have a certain *regard* for others, which we can now see consists in the capacity to take those others seriously, that is, either to treat them and their interests as having a certain worth or simply to be emotionally invested in their normative perspective, susceptible to respond emotionally to how they fare and how they are treated. To be accountable is also to be the appropriate object of confrontation, to be appropriately *held* accountable, so it also presupposes the capacity to secure uptake of the *to be communicated* emotional responses of anger and gratitude. This requires three more specific capacities: (a) the capacity to understand emotional communication (for example, to understand that the angry face one sees is an expression of anger and communicates a demand for acknowledgment); (b) the capacity to take up the normative perspective of others; and

[65] See Watson, *Agency and Answerability,* 219–59.

[66] Here is one place where I break, respectively, from McKenna and Watson, among others.

[67] For the charge that theories like Watson's are fundamentally forward-looking, see Matthew Talbert's "The Significance of Psychopathic Wrongdoing," forthcoming in *Being Amoral,* ed. Thomas Schramme, (MIT Press).

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

(c) the capacity to respond emotionally to slights and heights, such that one can feel what it was like for the slighted or heighted.[68]

IV. Conclusion

By acknowledging three distinct types of quality of will, each requiring distinct capacities and pointing to distinct conceptions of responsibility, we can make sense of our sometimes deeply ambivalent responsibility responses to marginal agents.[69] I will not go through all of the examples I mentioned earlier, but we can at least consider a few, starting with those Watson mentioned. Children *can* be malicious, and psychotics *can* be hostile. These are *aretaic* assessments, and so our associated emotional responses (such as disdain or abhorrence) track the quality of their character. Such responses are appropriate to the extent that these agents have (a) the relevant sort of character and traits, (b) the capacity for these traits to be expressed in their attitudes and actions, and (c) some sort of sustained agential identity. They might meet these conditions to some degree. But other sorts of responses, namely, disappointment, moral criticism, and varieties of anger, are not appropriate, precisely in virtue of these agents lacking the capacities for either quality of judgment or quality of regard. They *may* thus be attributability-responsible, but they are neither answerable nor accountable.

Psychopaths are more complex. They can indeed be indifferent to us, and they may judge that our interests have no worth. But this is not enough for them to have the capacities necessary for quality of regard: there is, after all, no communicative point in being angry at them, for they lack the ability to secure uptake of our anger, being unable to understand emotional communication or respond emotionally to our demand for acknowledgment.[70] They are not, on my schema, accountable for what they do. Nevertheless, they can be cruel, and they can exercise very poor judgment. In other words, they may be both attributability-responsible and answerable. And being responsible on both conceptions may appropriately generate the responsibility-responses characteristic of those conceptions, respectively, disdain and disappointment (as well as expressions of criticism).

Think next of those with advanced Alzheimer's disease. Their memory impairments and inability to engage in identifying empathy any more prevent them from securing uptake of the demands for acknowledgment

[68] Again, I am officially agnostic here on whether these conditions are together sufficient.

[69] I should also mention that, insofar as some of the conceptions of responsibility implicate multiple capacities, we may have ambivalent responses *within* these conceptions as well, where some capacities obtain but others do not. I lack space to discuss this possibility here, but I am grateful to Erin Kelly for drawing my attention to it.

[70] See also my "Psychopathy, Responsibility, and the Moral/Conventional Distinction," *Southern Journal of Philosophy* 49, Spindel Supplement (2011): 99–124.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press

built into anger and gratitude, rendering those demands pointless and those agents not accountable. They may also lack the capacity for rational judgment, rendering them no longer answerable for what they do. But they may nevertheless express in their actions remnants of longstanding character traits, such that, for example, admiration and other sorts of *aretaic* response are still warranted. They may, in other words, still be attributability-responsible.[71]

Finally, consider the clinically depressed. Their lack of accountability may trace to their inability to care about much of anything, so their ability to be emotionally pricked by what they do, or what happens, to others may be stunted. Nonetheless, their ability to make evaluative judgments may be left intact, even though they may lack deep attributability in a way, as earlier discussed. They could be examples of answerability without attributability or accountability. And their fractured responsibility grounds our fractured responses to them, perhaps borne out in our continued disappointment over their poor decisions but lack of disdain or anger in light of them.

My aim here has been to sketch out the tripartite view as a way to vindicate our ambivalent responses to a variety of real-life agents in a way that could blunt the motivation to go beyond facts about quality of will in developing a theory of responsibility. I have tried to do so by taking seriously the entire range of our responsibility responses—categorized by certain paradigm emotional pairings—and the way to avoid having to appeal to additional background beliefs is to develop a pluralistic account of their targets, an account about the three *qualities* of will and the three conceptions of responsibility they ground. This story, I think, while certainly more complex than any of the earlier, "purer" strains, more accurately captures who we are and what matters to us about agency.

Philosophy, Tulane University

[71] This is argued for at length in my "Remnants of Character," unpublished manuscript.

https://doi.org/10.1017/S0265052513000058 Published online by Cambridge University Press