Abstract: Two big ideas have taken center stage in recent discussions of moral responsibility: the idea that whether one is responsible, blameworthy or praiseworthy for an action is a matter of the quality of will manifested in the action, and the idea that it is instead a matter of what you do and whether it is in your control. These two ideas are often taken to be opposed to each other, appearing to give different verdicts in a range of cases from psychopaths’ crimes to expressions of implicit bias. In this paper I explore the nature of the opposition. In particular, I take up the question of whether proponents of the two groups are sometimes talking past each other by aiming to explicate distinct concepts (for example, to oversimply, one group is interested in what it takes to be deserving of some harm or benefit while another is instead more exclusively focused on what it takes for certain moral emotions to be appropriate). In working out the answer to this question, I show how we are led to the more fundamental question of whether we can or should separate debates about desert from those about the aptness of moral emotions, appropriate changes to relationships and more.

1. Introduction

Many working on the nature of moral responsibility have divided themselves into two camps which I will call “quality of will theorists” and “control theorists”, respectively. Quality of will theorists, as I will understand them, are a large and varied group, and include those who take it to be sufficient for being responsible and blameworthy that one has acted from a bad quality of will, or has manifested bad values in one’s judgments. Control theorists on the other hand, take it that one must (also) have had some control over what one is blameworthy for, and this often takes the form of one’s having been able—in some relevant sense—to have had the opportunity to have recognized and acted for different and better reasons at the time of action.

This paper is inspired by the observation that along one dimension, quality of will theorists (or at least some prominent ones) tend to go in for relatively more blaming than control theorists in their verdicts about blameworthiness while along another dimension, they tend to go in for something more minimal. More specifically, although quality of will theorists tend to be more inclusive than control theorists in what they count within the scope of the blameworthy, and so on this dimension seem to be in the “harsher” camp, they also tend to connect blameworthiness less often to ideas of desert or sanction, and to the extent that they see those latter ideas as parts of quite different discussions altogether, they seem the “kinder and gentler” group. Could it be that the two camps are talking past each other? What explains...
these different combinations of commitments? And where should the conversation go from here?

Before going further, it’s important to make a few preliminary points about how I’ll approach these questions.

First, I am going to set aside at least for the most part what some might see as the elephant in the room, namely, the debate about whether indeterminism is also required for responsible action of any kind. This is an important question, and it might turn out to be essential in the end for adjudicating the debate between these two camps if the conditions that control theorists appeal to are ultimately best understood as indeterministic. But interestingly, much of this debate between quality of will theorists and control theorists takes place between different kinds of compatibilists about determinism and responsible agency, or at least between those who have set aside the question of whether determinism is required for the requisite control. I believe that progress can be made, then, by at least first focusing on the debate between these theorists.

Second, especially in the first half of the paper, I will be painting in fairly broad brush strokes (which is a bit outside of my normal approach). So I want to acknowledge at the start that there are notable exceptions to the broad generalizations I make and that, a fortiori, I do not see them as necessary. At the same time, I think that we can gain insight by exploring why they might hold in some systematic if nevertheless contingent way. Relatedly, it is also important to note that there are views that do not necessarily fall neatly into either of these camps. For example, some suggest as a requirement for blameworthiness the possibility of uptake in moral address that might go beyond a condition related to the will of the agent, but isn’t yet to require control in a robust sense. While these are important views and deserve their own treatment, for the purpose of this paper I focus primarily on the two camps mentioned earlier.

Third, in addressing the question of whether theorists are talking past each other, in what follows, I’ll make use of a methodological distinction made by John Rawls in A Theory of Justice. Rawls there distinguishes between the concept of justice, and competing conceptions of justice. There is a concept of justice which theorists agree about—in Rawls’ words, “that institutions are just when no arbitrary distinctions are made between persons in the assignment of basic rights and duties and when the rules determine a proper balance of competing claims to the advantages of social life” (1971, p. 5) Theorists then disagree about all sorts of things—what is the proper balance of competing claims? An equal distribution? Is it a utility-maximizing one? A libertarian one? These answers all capture competing conceptions of justice. Without doing Rawls exegesis here, I want to focus on an important distinction inspired

---

2 See, for example, Watson (2011) and MacNamara (2015). McKenna (2012) also adopts a view of this sort, while combining it with others. Wallace (1994) might be another example, among others, of a view that doesn’t fit neatly into either category.
by Rawls’ discussion that can be drawn as follows: the concept is the notion that we all agree about and are trying to theorize about. Competing conceptions are attempts to identify the more specific conditions (often, but not always, necessary and sufficient) for the instantiation of the concept. In the case of justice, the questions are what counts as arbitrary differences among people, and what counts as the proper distribution of advantage. A natural question in the context of the debate about responsibility, then, is whether different parties are focused on the same or different concepts.

Because one of the questions on which I focus is precisely whether we are talking past each other, I will try for as long as possible not to elucidate further what is meant by “responsibility” in the debate.

In the next section, I set out in more detail how I am understanding the views in the two camps, and offer some examples of their verdicts in particular cases. In sections 3 and 4, I consider hypotheses regarding the location of the ultimate disagreement between the views. In section 5, I offer a path to ensuring greater direct and substantive engagement between the two camps that rests on a particular understanding of desert, and in sections 6, 7 and 8, I consider potential challenges for taking this particular path before concluding in section 9.

2. Quality of Will vs. Control: The Accounts and Their Verdicts

Each camp inhabits a large tent, to continue the metaphor—that is, there is great variation among members of each group. Very broadly, I see the contrast as follows: quality of will theorists take it that one is blameworthy when one’s actions or omissions or mental states manifest ill will or indifference where good will is called for; control theorists take it that one is blameworthy when one’s actions or omissions are in some sense in your control, where this typically assumes that there is a sense in which you could do otherwise, and that you could exercise normative reasons-responsive capacities in your possession.

It will be helpful to provide a few illustrations of views that fall into the quality of will camp. Matt Talbert, summing up the views of “attributionists”, whom he takes to be a subset of Quality of Will views writes,

The approach can be glossed this way: assessments of moral responsibility are, and ought to be, concerned with what is attributable to an agent in such a way that she would be eligible for moral praise or blame on the basis of these attributions, and the attributions that warrant such responses are those that implicate morally significant features of an agent’s orientation toward others.” (Talbert 2022 bold emphasis mine)

T.M. Scanlon writes that blame (in a variety of forms) is made appropriate “by the fact that a person’s attitudes are deficient as measured by the standards of morality or some relevant interpersonal relationship.” (Scanlon 2013, p. 106.) And Angela Smith writes that “accountability blame” is appropriate “whenever an agent’s conduct manifests unjustified ill-will or disregard toward others” (Smith 2019, p. 107).
In contrast—and to put my cards on the table—on the control view I favor, to be blameworthy, one must act (or omit to act) wrongly with an opportunity of sufficiently high quality. The quality of an opportunity is a function of the nature of the agent’s capacities, or their normative competence, and situational factors that enhance or detract from the situation’s congeniality to acting well or badly.³

While there are many interesting and important differences between theorists within each camp, for present purposes, I will set those to the side. I think we already have enough to see that there are important categories of cases in which views in the two camps appear to deliver different verdicts.⁴

Perhaps the most high-profile category of cases thought to divide the two camps are psychopaths—typically understood to be those entirely lacking the capacity to grasp moral reasons. So understood, psychopaths fail to fulfill the requisite conditions for blameworthiness on most control views, while it is claimed that they will meet those conditions on most quality of will views.⁵ One might question (as I have elsewhere) whether psychopaths meet even the minimal conditions for expressing a bad quality of will such as manifesting cruelty, given the stipulation that they are unable even to grasp (let alone reject in a particular case) such propositions as that others’ interests are reasons for acting.⁶ If so, they could not serve as a test case. However, even if this skepticism is borne out, it seems that there are cases in the ballpark, perhaps those we might think of as having “selective” psychopathy and who are simply unable to recognize the salience of moral reasons in certain fixed contexts. For example, one kind of case includes people who are so-called “incurrigible” racists (which there are empirical reasons to doubt exist in any significant numbers today especially)⁷. As with other cases I describe here, it is crucial to describe the cases fully, and in a way that ensures that readers accept the relevant stipulations if they are to serve as genuine test cases. This is not always easy to do, and intuitions can shift once people have the stipulations properly in view.⁹ But once we do so, in these cases, quality of will views take the protagonists to be blameworthy, while control views do not.

Responsibility and blameworthiness for implicit or unconscious attitudes and their manifestations are a second sort of case that has sometimes been thought to divide the views.

³ Other varieties of control views can be found in bodies of work by Susan Wolf, Randolph Clarke, David Brink, and Manuel Vargas, and John Martin Fischer and Mark Ravizza, to name a few. However, I should note that not only can it sometimes be the case that views do not fall into one or the other camp (because there are additional camps), but it can sometimes be difficult to be sure whether a view falls squarely into one of the camps or not.
⁴ Unless, that is, we are able to locate a notion of quality of will that is a function of what you do with control—something I am also interested in pursuing. See Nelkin (2016).
⁵ See, for example, Brink (2013) for the first point, and Talbert (2008) and Smith (2019) on the second, respectively. It is noteworthy that Brink goes on to question whether actual diagnosed psychopaths have the deficit in question. Nelkin (2015).
⁷ See Watson (2011).
⁸ See Ryazanov et al. (2018) for studies showing that people systematically override stipulated features of at least some thought experiments with their own prior background beliefs.
Consider cases in which the attitudes in question are paradigmatic examples of bad quality of will, such as the racist or sexist attitudes that members of some group are not worthy of the same respect as those of another group. If one’s attitudes are truly inaccessible to one, and yet they are efficacious in producing action, it seems that at the moment of action, one lacks the relevant control or opportunity to avoid so acting. On first glance, then, control views would appear to excuse these individuals, in contrast to quality of will views which can point to the ill will manifested in the content of the attitudes in support of a judgment of blameworthiness.\(^{10}\)

This is a particularly tricky kind of case for a variety of reasons. One is that in real life cases, there are often earlier moments in which people have control and could have done things to prevent their attitudes from manifesting. Many control views accept a so-called “tracing” condition, allowing for indirect responsibility for actions by virtue of such an earlier moment of control. Thus, in many real life cases, control views might join quality of will views in holding people who act badly while manifesting implicit attitudes blameworthy.\(^ {11}\) Similarly, there is room for some quality of will views to withhold judgments of blameworthiness if, say, the implicit attitudes in question do not reflect the true quality of a person’s will. But such cases have certainly been taken to be paradigm test cases to the extent that they can be properly described as revealing quality of will while lacking a kind of control. So, again, without getting into the details of the cases, I put this forward as a category that has at least been thought to be a test case.\(^ {12}\)

A third kind of well-known case that appears to offer a test for comparing the two kinds of accounts is that of the “willing addict” who acts wrongly. Frankfurt (1971) presents a contrast between someone who is addicted to a substance and who strongly desires that he is not, and someone who is similarly addicted, but wholeheartedly endorses the addiction and the behaviors that flow from it. In a recent and subtle paper, Chandra Sripada (2017) elaborates on the details of these cases and claims that refined versions of them provide one of the few true tests to help us decide between control views and quality of will views. I agree that when fully described—a subtle matter that includes, among other things, stipulating that the individuals are not themselves responsible for their current situations by virtue of tracing to some earlier actions or omissions when they did have control—such addiction cases can serve as test cases in the evaluation of the debate. Here, too, quality of will views include the willingly addicted person who acts wrongly in the group of blameworthy agents, while control views excuse.\(^ {13}\)

---

10 See, for example, Smith (2005), p. 256.
11 See, for example, Fischer and Tognazzini (2009) and Nelkin and Rickless (2017). Some control views do not require awareness of salient aspects of the situation for the relevant control, as long as one could have had such awareness. See Clarke (2014), for example. On this kind of view, tracing would not necessarily be needed to include such cases as blameworthy. So this case is also a good test case to distinguish competing conceptions under the control tent.
12 See Holroyd (2012) and Levy (2017) for discussion of these cases in the context of assessing responsibility with respect to differing accounts of the conditions for blameworthiness.
13 See Koepke (manuscript) for a response to Sripada that shows that it is more difficult than presented to describe the cases in ways that clearly satisfy the quality of will condition, but not the control condition.
Finally, a less often discussed set of cases is a fascinating, but particularly horrible, form of dementia, on which cognitive functioning remains largely intact at the early stages, while an impairment of the ability to care about others seems to arise. People with the diagnosis of fronto-temporal dementia have been known to act in very hurtful ways to those with whom they were extremely close. On control views, this case would appear to be one in which the agent is not blameworthy, while as I’ve argued, notable quality of will views would appear to imply blameworthiness. (See Nelkin (2019)).

These are just a few kinds of cases in which verdicts from the two camps have been thought to diverge. In general, it appears that the quality of will theory has a more expansive scope. Thus, much hinges on adjudicating this debate, since the differences are potentially significant. If we focus only on the extension here, the quality of will view appears the harsher.

And yet, while control theorists often discuss their views in terms of desert and sanction, it is striking that this is not the case for many quality of will theorists. Some anecdotal evidence: a search of Smith’s papers reveals almost no use of the term and we sometimes find that the idea of desert is explicitly divorced from the topic at hand (e.g., Shoemaker (2015), Scanlon (1986) and (2008) and Hieronymi (forthcoming). This can make it seem like it is control theorists who are the harsher group. And indeed this is the interesting combination of features that initially prompted me to wonder whether we are talking past each other. But first, let me emphasize an important caveat mentioned in the introduction. Some views that might be classified as quality of will views, at least when very broadly understood, countenance desert. For example, so-called “character” views, much discussed in American jurisprudence, are a case in point. It is true that they may not be central cases of quality of will views, even on our expansive understanding, in so far as they appeal to character which is a state of agents over time, and they are fundamentally about criminal, not moral responsibility. But I do note that they share with quality of will views a rejection of the importance of control or abilities, even general ones, to do otherwise. In any case, the more important point is that the association I have been pointing to between quality of will views and a recoiling at thoughts of desert and sanction is a contingent and imperfect one. On the flip side, the idea of desert is not always center stage for control theorists either. Still, if there is some reason for a systematic, albeit imperfect divergence on this point, that itself would be interesting.

---

14 One exception is Smith (2019) in which she argues against the position that the negative reactive attitudes have as part of their content that their objects are deserving of suffering. (Her main target here is Rosen (2015).) While this does constitute engagement with the notion of desert, but only to rebut the idea that it is relevant. I address her reasoning in section 6.

15 It is worth noting that Scanlon, in rejecting any “pre-institutional” notion of desert (1986, p. 188-89), takes desert to have implications (such as that if someone deserves to be treated harshly, it is good that they are so treated) that I believe we can reject without rejecting a role for desert (p. 172). I return to this point in section 5. It is also important that he “is content to be offering a revisionist theory” if a compatibilist account must accept implications of blameworthiness for desert (p. 172). And similarly, in (2008), Scanlon acknowledges that the account he gives there might be revisionist.

16 See, for example, Moore (2010) for a comparison of “character” and “choice” views.

17 For example, it does not feature centrally in much of John Martin Fischer’s work defending one of the most influential control accounts of responsibility. However, it can be found: in offering a summary of his “semi-
To return to our first descriptive question about the debate, then: Are the two camps really trying to offer conditions for different concepts?

3. A First Answer: Two Concepts, Namely, Attributability and Accountability

There has been much discussion of different concepts of responsibility, particularly since Gary Watson’s highly influential “Two Faces of Responsibility” in which he distinguishes between responsibility as attributability and responsibility as accountability. Following Watson (to start) we can distinguish between two notions of responsibility. According to Watson, one is responsible in the attributability sense if one’s actions reflect one’s having adopted an end, one’s having committed oneself to a certain conception of value. To blame someone in this sense is to attribute a moral fault to an agent (Watson 2004, p. 266). In contrast, one is responsible in the accountability sense if it is appropriate to make moral demands of one as an agent. And our practices of holding one another accountable ‘involve the imposition of demands on people’ (Watson 2004, p. 273). In turn, according to Watson, imposing demands is a matter of laying it down that ‘unless the agent so behaves she will be liable to certain adverse or unwelcome treatment’ or sanctions (Watson 2004, p. 275). Sanctions are in turn connected with reactive attitudes such as indignation, resentment, and disapprobation in the following way: either they are themselves disagreeable when experienced by their targets or they involve dispositions to treat others in generally unwelcome ways. Because the prospect of adverse treatment arises in this way, questions of fairness arise in connection with accountability, and it is for this reason that many have thought being able to do otherwise is essential for accountability. In order to be blameworthy in the accountability sense, the demands in question must be fair or just.

Now, as has been noted, there is a lot to unpack in the characterization of each of these notions. We will return to this point, but we can already see that it is plausible that there are two distinct concepts. It would be neat if quality of will theory mapped onto the former and control onto the latter, and Neil Levy and Michael McKenna, among others, speculated not long after this distinction entered the conversation, that this is precisely what is going on. (Levy and McKenna (2005); McKenna (2012).) We would have two different concepts, each with a different—possibly correct—conception. Quality of will views seem to go nicely with the idea of one’s actions manifesting the adoption of (good or bad) ends, and with the idea of attributing faults or virtues, such as that one is cruel or generous. Control views, on the other hand, seem to go nicely with the idea of one being subject to demands and thereby acquiring a certain liability to sanction.

compatibilism” in Vargas, Fischer, Kane, and Pereboom (2007), he describes it as an advantage of his conception that accepting it does not require us to “etiolate or reconfigure the widespread and natural idea that individuals morally deserve to be treated harshly in certain circumstances, and kindly in others. (p. 82, emphasis his). Thanks to Derk Pereboom for this reference.
Alas, this is too neat. One superficial reason for this is simply that theorists have denied precisely those characterizations of their own work. Angela Smith, for example, writes: "we shouldn’t be “misled” into thinking that apparent disagreements aren’t really disagreements." (Smith 2015, p. 110, p. 122). But a deeper reason is that they seem right about their own views! Most centrally, the negative reactive attitudes often play a large role in the introduction of quality of will views, a role that in some cases is used to fix the concept of responsibility in question. So it cannot be that quality of will conceptions are just trying to capture attributability understood either as an as an attribution of good or bad ends, or as a judgment of vice or virtue, in opposition to accountability, which is thought to make the reactive attitudes appropriate.

4. One Concept or Two?

This thought leads naturally to the idea that we are all talking about the same concept after all, one for which the reactive attitudes play the central role, and, in particular that the concept is that of accountability.

Interestingly, members of both camps often point to the classic source of the label “reactive attitudes”, namely, Peter Strawson’s (1963/2003) “Freedom and Resentment” for a way of fixing the concept at issue. They sign on to some variant of the following biconditionals:

(The Reactive Attitudes-Responsibility Biconditional) X is responsible for A if and only if it is appropriate to respond to X with some subset of the reactive attitudes.

And

(The Reactive Attitudes-Blameworthiness Biconditional) X is blameworthy for A if and only if it is appropriate to respond to X with negative reactive attitudes such as resentment, guilt, and indignation.

Strawson’s article has also been a touchstone for contemporary quality of will views, since he claims there that the reactive attitudes are fundamentally responses to the quality of others’ wills toward us ((1963/2003, p. 76). At the same time, we find that Strawson also includes reference to desert and sanction. As he introduces his subject, he writes that he will set aside punishment and condemnation “at first” to focus on the reactive attitudes (emphasis mine). He later continues,

“Only by attending to this range of attitudes can we recover the fact as we know them a sense of what we mean, i.e., a sense of all we mean when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice (91).

Whatever else Strawson is doing in the article, he is clearly thinking that all of these ideas are all part of the same conversation. This suggests that it should be a constraint on our theorizing about blameworthiness and responsibility that we accommodate at least the
appearance that these concepts are connected in important ways. If we were ever in need of an example to show that the separation some quality of will theorists make between their account of blameworthiness and the notions of desert and sanction is a contingent one, Strawson appears to provide it.

But the authors I have quoted above as defending quality of will views do not take the conversation to move quite so smoothly as Strawson seems to have from reactive attitudes to desert and sanction—at least when “desert” is taken to mean something beyond “appropriateness” or “fittingness” and to apply to anything other than the attitudes. It is important to distinguish desert in the context of debates about moral responsibility from mere appropriateness, which is a category that is correctly used to describe many kind of emotions and attitudes. Consider, for example, the emotion of grief. It is appropriate when one has lost a loved one, and part of what makes it appropriate is that it accurately reflects important facts about the world. Yet we would not typically say that one deserves to feel grief. “Desert” as it traditionally appears in debates about free will and moral responsibility goes beyond this notion of “appropriateness” that clearly applies to grief among other emotions and attitudes.

It is tricky to say more positively what desert is, and I am sympathetic to Joel Feinberg’s suggestion that it is not analyzable (1970). Not being analyzable does not entail that we cannot say things about it, however. And I believe we can. For example, what is deserved is valenced and scalar in a way parallel to the objects of desert, that is, to what it is that renders one deserving. And I take it that I take it that in the relevant sense the modes of treatment that are deserved are those that either set back or advance a person’s interests. There is much debate about what else is true about desert, but among the most well-known claims are that when a people deserve something, it is good that they get it, or that it is connected to reasons we have for giving people what they deserve. I will return to these suggestions in the next section. For now, it is important to note that desert, as I understand it here, is not to be understood simply as appropriateness. And it is in this more robust sense of desert that at least some quality of will theorists take desert to be part of a different conversation altogether. It is also important that a person does not deserve something in the relevant sense simply in virtue of its having a good consequence or in virtue of its being something that results from a set of rules we have agreed to. Derk Pereboom identifies a notion of responsibility that is essentially connected to what he calls “basic desert” and that captures these features:

For an agent to be morally responsible for an action in the basic desert sense is for the action to be hers in such a way that she would deserve pain or harm if she was sensitive to its being morally wrong, and pleasure or benefit if she was sensitive to its being morally exemplary. The desert at issue here is basic in the sense that the agent, to be morally responsible, would deserve the pain or harm, the pleasure or benefit, just by virtue of having performed the action with sensitivity to its moral status, and not, for example, by virtue of consequentialist or contractualist considerations. (2022, p. xx)\textsuperscript{18}

\textsuperscript{18}Pereboom’s characterization has changed from those earlier (such as in (2001) and (2014)) in a notable way. In particular, this new characterization explicitly accommodates the fact that the notion of desert in this context is
Now it is a very important question what exactly the relationship between desert and accountability is. At the same time, it is also clear that while many control theorists, alongside Strawson and Watson, take them to be connected, influential contemporary quality of will theorists often avoid or even balk at making a connection. For example, David Shoemaker write, at the very end his book, *Responsibility From the Margins* (2015): “We may begin by noting the obvious: I have said nothing whatsoever about desert, condemnation, or justice” (p. 220). (He also notes that he has not mentioned free will.) That is a somewhat remarkable thing to note after a book-length treatment of responsibility and a defense of a quality of will view of three distinct notions of responsibility, including accountability. And yet Shoemaker seems to take it that the question of whether someone is blameworthy or not, or what blameworthiness amounts to, is not to be answered even in part by considering desert or sanction or condemnation. There is a further moral question about whether sanction or punishment or reward is appropriate all things considered, according to Shoemaker, but this is really a different question, and one that isn’t even special to responsibility cases. I believe almost everyone would agree that whether a response—even just expressed blame—is appropriate all things considered is not fully answered by whether someone is blameworthy. But Shoemaker seems to go beyond this thought in not taking desert to be implicated by blameworthiness. And indeed he takes the division between those who see desert as essentially connected to responsible agency and those who do not to be a major “fault line” in the debates about moral responsibility and that on either side lie such radically different approaches that “it is unclear how they may ultimately be brought together.” (2020, pp. 221-225.)

If the implication of this approach is that we ought to decide on a conception of blameworthiness without thinking about these further moral questions, then it raises the question whether the concept of blameworthiness Shoemaker and others are working with is just a different concept than that of the control theorists. Pereboom, who takes it that basic desert has been central to traditional debates about free will and moral responsibility is skeptical that it is ever instantiated. And yet, he has defended in great detail the existence of a distinct kind of responsible agency, where “responsible” is understood in a different sense. It is a notion of responsible agency that makes protest appropriate, and notably for present purposes, he cites with approval a number of quality of will theorists for having offered persuasive accounts of blame as a kind of protest that can be appropriate without the targets of blame having had control.\(^\text{19}\)

Thus, there is some reason to believe that the concept that Shoemaker and others are working with is really quite different from one that seems essentially connected to these others.

\(^{19}\) See Pereboom (2017) and (2022). It should be noted that Pereboom takes the reactive attitudes such as resentment and indignation to be inappropriate if no one is deserving in a basic sense, and this marks a serious disagreement between him and quality of will theorists who take themselves to be giving a conception of moral responsibility that entails the appropriateness of such attitudes.

understand to require that what is deserved be either setbacks of interests (in his words, “harm or pain”) or advancements of interests (“pleasure or benefit”).
If this is right, then perhaps the disagreement is really at a higher level, the level of concepts. Perhaps we really are aiming at two different concepts. On this view, at least one party takes it that the concept in which the other party is interested is simply incoherent or, at the least, uninstantiated. But this is consistent with our being engaged in a common enterprise. We might be trying to fix the relevant concepts in similar or overlapping ways. Everyone identifies a concept that they believe answers to the reactive attitudes in some important way as well as to some central cases. But for at least some proponents of influential quality of will views, this property is seen as quite separate from anything having to do with liability to sanction or desert or free will, whereas for at least many control theorists, the property picked out is one that is essentially connected to these other concepts. Even on this view, there is a way in which we agree on a lot. And this makes sense—we all want to capture that phenomenon or set of phenomena that seems so important in human life that Strawson thought we wouldn’t really be human without it.

It is equally important to take note of an interesting overlap of commitments that can be obscured when we focus on the treatments of various case studies by the different accounts. It is a commitment not to attribute a robust notion of desert to those with significant control-related impairments as exemplified by (properly described) psychopaths or their more selectively impaired cousins, those who lack control due to the operation of unconscious attitudes, and many others. There are at least three ways to avoid such attributions: set a higher bar for responsibility and blameworthiness in the way that control theorists do; shave off a robust notion of desert from blameworthiness across the board; or both. Control theorists tend do the first, while at least some quality of will theorists do the second. Each in their own way issues less harsh verdicts about a host of cases than seems to have been the norm until relatively recently, and they converge in doing so.

This provides at least a partial answer to our second question by offering a partial explanation of why the different sets of commitments of each camp (contingently) fit together.\textsuperscript{20} They provide different ways of fitting with particular verdicts that strike many as appealing. At the same time, each has an understandable complaint about the route the other has taken to these verdicts, as well as its further implications and additional verdicts. Control theorists worry that quality of will theorists have simply left out a real part of the picture by opting for a notion of responsibility that does not bear a straightforward connection to the notion of desert. Quality of will theorists worry that control theorists have left us without strong enough appropriate responses to genuinely malicious or insidious wrongful behavior.

In the end, I have provided some reason to believe that members of the two camps are not all quite on the same page when it comes to the concepts we are aiming to elucidate. But I

\textsuperscript{20} A fuller explanation will likely require getting into specific details of the views. For example, Shoemaker combines commitments related to desert with other commitments, including a non-cognitive conception of the reactive attitudes and a response dependent account of responsibility. (See, for example, Shoemaker 2018 for a defense of these other commitments, and see note 23 for brief discussion) And Scanlon and others who adopt a contractualist view of normative ethics have distinctive resources for generating conclusions about desert.
leave the answer to the descriptive question somewhat open. Perhaps appearances are misleading.\textsuperscript{21} Whether we already \textit{are} talking more directly to each other or not than the “two concepts” answer entails, in the next section, I propose a path for ensuring more direct engagement.

5. Where To Go From Here: A Proposed Path Toward a Common Concept

The basic idea behind my proposal is to identify a substantive notion of desert that accounts for the idea that our conclusions about moral responsibility have implications for discussions of desert, justice, condemnation, and even perhaps punishment, as Strawson eloquently articulated. At the same time, the notion need not be burdened by all of the typical commitments of self-described retributionists.

I propose this intuition pump: someone needs to be sanctioned in order to deter further massive violence, and there are two candidates—a person who was responsible and blameworthy for committing a crime or someone else who has an excuse. Is there really nothing to choose between here? (And to fill in the case further, imagine that we could \textit{communicate} our feelings to the guilty party even they are not the one sanctioned.) If the answer is yes, then it seems to me that there is a tight connection between blameworthiness and sanction. And arguably the connection is desert. Being deserving makes one \textit{liable} to sanction. It provides a conditional reason for giving someone what they deserve. We can capture the basic idea as follows:

(\textit{CR}) X’s being deserving of sanction is a conditional reason to sanction X. (Nelkin 2016)

Note that accepting this point does not require accepting theses often associated with desert: there is no need to say it is good for people to get what they deserve or that we ought to give people what they deserve in the absence of other considerations; it simply shows that blameworthiness entails a (weak) \textit{liability} to sanction.

It is true that many control theorists accept some of these additional theses, such as those who advocate the goodness of blameworthy people of getting what they deserve or that there is always \textit{pro tanto} reason to provide it, but it is possible to be agnostic about or reject these additional points and yet accept (CR).\textsuperscript{22} If so, I believe that we have a way to capture

\textsuperscript{21} As one example, I note that although Smith takes sanctions to be a legal or divine matter, she acknowledges that the conditions for responsible agency she identifies “may well be among the necessary conditions for the just distribution of such benefits and burdens” (2015, p. 106). Although presupposing that sanction isn’t itself a moral notion, it suggests some room for a connection between blameworthiness and sanction. Similarly, while Shoemaker cites Elinor Mason’s (2019) recent treatment of blameworthiness as on the side of the fault line that eschews desert, I note that Mason seems to be concerned to avoid the issue and not to argue positively that desert has no role to play. She writes that she does not “frame the issue [of blameworthiness] in terms of desert,” and instead provides an account of when it is fitting to blame. But she also takes her account to be “consistent with various views of desert” (p. 7).

\textsuperscript{22} In Nelkin (2016), I argue that we should in fact reject these further theses. See also, among others, Feinberg (1970b) for a defense of the thesis he calls, “Fault Forfeits First”.

12
desert as essentially connected to a potentially shared concept of responsibility. I believe that intuition pumps like the one offered above support the acceptance of a notion of desert invoked by (CR) by revealing an explanatory role. Further, accepting this connection also allows us to preserve what until relatively recently seemed to be frequently presupposed, namely, that questions of desert and even punishment were tightly related to questions of moral responsibility.\(^{23}\) And it goes beyond an adoption of the label of “desert” as meaning “appropriate” in a way that grief is appropriate as a response to loss. This notion connects responsibility to justice.\(^{24}\)

Now on this way of locating a shared subject matter, it would be inadvisable to completely ignore questions concerning the treatment of others that go beyond the minimal immediate expression of the reactive attitudes, or of a justified distancing, in identifying the correct conception of responsibility and blameworthiness. This might require a shift in methodology on the part of some of the quality of will theorists I’ve mentioned. At the same time, it would have the benefit of allowing us a way of seeing us as talking more directly to each other.

Not surprisingly, there are ways to challenge this approach. One is to reject the project of seeking a shared subject matter altogether and instead recognize two separate areas of study.\(^{25}\) While I understand the temptation to do so, I believe that the advantages of identifying a shared concept are so significant that we should only adopt this bifurcated approach if attempts to unify the conversation fail. In what follows, I briefly canvass and reply to three objections to the particular attempt at unification just sketched.

6. **First Challenge: The Reactive Attitudes Biconditionals Capture All We Need**

In this section I would like to focus on one argument that proponents of quality of will views offer in support of taking desert out of the picture. It is to focus on the biconditional of blameworthiness, and then to show that there is nothing about the reactive attitudes that

---

\(^{23}\) I have also argued that being deserving in this sense has the same satisfaction conditions as being responsible (namely, a control conception of responsibility) (Nelkin 2016.) This not only provides an argument connecting the two concepts but it also offers a specific way that they are connected. I do not appeal to that argument here, as it might appear question-begging in the context of examining a debate between those advocating different conceptions of responsibility. But I do note that the various theses provide mutual support insofar as they hang together.

\(^{24}\) Mitchell Berman (2021) has recently argued that while liability to sanction attaches to blameworthiness, desert, being a “favoring” relation, is an entirely separate notion. Here, in contrast, I reject the idea that desert is a favoring relation, and so take both liability and desert (distinct concepts to be sure) to attach to blameworthiness. Berman’s argument is a rich one, and a full evaluation must await another occasion. For now, I just focus on one foundational element: the claim that our concepts of desert and blameworthiness have distinct “functions”. In reply, if we take concepts to come apart in the sense that they lack entailments when they have different functions, the reasoning would prove too much. For example, it could be argued that a number of the different sub-species of blame identified by Berman would themselves then seem to be conceptually unrelated.

implicates desert. (See Smith (2019) and Shoemaker (2020) for two examples.) Here is a version of the argument that I believe brings out key assumptions:

(1) (The Reactive Attitudes-Blameworthiness Biconditional) $X$ is blameworthy for $A$ if and only if it is appropriate to respond to $X$ with negative reactive attitudes such as resentment, guilt, and indignation.

(2) If the negative reactive attitudes are not themselves necessarily harmful and their content does not include that anything is deserved in response, then the reactive attitudes can be appropriate without anything harmful being deserved.

(3) The negative reactive attitudes are not themselves harmful.

(4) The negative reactive attitudes do not have content that includes that anything is deserved in response.

Therefore,

(5) The reactive attitudes can be appropriate without anything harmful being deserved.

Therefore,

(6) One can be blameworthy in the accountability sense without desert.\textsuperscript{26}

I believe that this argument, while valid, is unsound. In particular, we have good reason to question premises (2) and (4). Let us begin with (2). Some (e.g., Gideon Rosen (2015)) have argued that it is a constitutive thought of resentment that suffering is deserved. If this were correct, then the reactive attitudes would be inextricably linked to desert, and it would not be possible to disassociate them. (2) is designed to rule out this kind of linkage (among others). But it is possible to take resentment and other negative reactive attitudes to implicate desert without it’s being part of the content and without it’s being consciously “thought of”. For example, Randolph Clarke (2016) takes it that guilt requires the belief that one is blameworthy.\textsuperscript{27} This leaves it as an open question whether that entails a commitment to desert. Thus, the reactive attitudes and desert might be inextricably linked in additional ways, including in the case that the attitudes brought with them a kind of cognitive commitment that is weaker than occurrent belief. This sort of commitment would also have to be ruled out in order to ensure that the reactive attitudes can be appropriate in the absence of anything being deserved. Thus, a revised premise (2) is more plausible:

(2’) If the negative reactive attitudes are not themselves necessarily harmful and their content does not include or entail that anything is deserved in response, then the reactive attitudes can be appropriate without anything harmful being deserved.

\textsuperscript{26} This argument is a version of reasoning that I find in Smith (2019) and Shoemaker (2020). Smith, who offers the more detailed version, presents her argument as a rebuttal to an argument that concludes that desert is relevant to accountability (and further, that control is also required).

\textsuperscript{27} I endorse something like this approach in Nelkin (2016) and in Brink and Nelkin (forthcoming).
This means that for the argument to remain valid, we also need a revised premise (4):

(4’) The negative reactive attitudes do not have content that includes or entails that anything is deserved in response.

But now it is easier to see how the premises can be questioned. When it comes to premise (4’), it is not enough to point to cases in which one feels resentment but has no desire that the object suffers or has no belief that it would be good that they suffer. We might initially think that a claim like (4) is supported by the following sort of case: one can feel indignant toward one of one’s children for hurting the other without wishing any harm on either or thinking that harm of any kind would be a good thing. Similarly, Smith (2019) appeals to an example of resenting her colleagues for not doing their fair share of department work but in which she wishes them no harm or has any thoughts of desert. But this sort of case, while extremely plausible, does not support (4’) once we have the correct understanding of it. One reason is simply that (4’) does not require that one consciously think a thought of the sort that it would be good for someone to get what they deserve, which we can see now that we have revised the premise. But an even more important reason, in my view, is that, with the correct account of desert in hand, it is clear that one can consistently believe or simply be committed to its being the case that that someone deserves something of a setback to their interests (which might not amount to felt suffering at all), without believing or even being committed to the thought that it would be good for them to receive it. Thus, the kinds of cases sometimes put forward to show that we can resent without implicating desert are insufficient by themselves. These cases fail to show that desert is not implicated if we have the right understanding of desert.28

I also believe that premise (2’) can be questioned. It rests on the idea that if the blameworthiness biconditional is true and desert is related to blameworthiness, then desert and the reactive attitudes must be directly connected to each other. But this idea might be an artifact of an overemphasis on the reactive attitudes biconditionals. If one focuses exclusively on them, then it can seem that everything related to responsibility must go through the reactive attitudes. But just because they give us one set of necessary and sufficient conditions for responsibility and blameworthiness does not mean that there are not others. This could be the case if the same further condition—say, having a high enough quality opportunity to act well instead of wrongly—provided the right conception for both desert and the

28 As I understand it, Shoemaker (2018) and (2020) adopts a position of non-cognitivism about the emotions, including the reactive attitudes, suggesting that they do not require particular cognitive content. In that case, having the attitudes in question would not commit someone either to a belief or to a commitment to any truth regarding desert. An additional argument would be required to rebut a defense of (4) or (4’) that rests on such a theory, but I lack space to include this here.

29 For an argument that the very content of the attitude of guilt in particular contains the self-referential idea that the unpleasantness of the very attitude of guilt is deserved and that it is a better world in which people get what they deserve, see Portmore (2019). I believe that there is good reason to question both of these claims, though lack the space here to pursue them.
appropriateness of the reactive attitudes. In fact, I also believe that the reactive attitudes and desert are more directly connected, as well. But it is worth stepping back for a moment and asking whether they must be.\footnote{For related discussion of this issue, see Nelkin (2013.).}

When we put these points together, there are multiple reasons to believe that accepting the reactive attitudes biconditionals is compatible with seeing a deep and essential connection between blameworthiness and desert.

7. The Second Challenge: A Tension Between Responsibility as a Cross-Domain Concept and an Essential Connection with Liability

A second challenge for the approach I have suggested is that moral responsibility seems to be but one kind of responsibility among others; there seem to be other domains of responsibility and blameworthiness, including the aesthetic, the epistemic, and the athletic, among others. If this is right—and it’s controversial, to say the least—then there is pressure to accept that the concept of responsibility might be more “austere” to use a phrase Matt Talbert uses to describe a particular conception of responsibility. In this case, perhaps we ought to see desert as entirely separate from responsibility. Does responsibility in the aesthetic or epistemic realm—when divorced from any moral considerations—really commit us to the idea of desert or liability to sanction? If I paint or play basketball badly, am I liable to sanction? If we are to preserve the parallel of responsibility across domains—and I’ve argued for this myself—then it might seem that we should go more, rather than less austere.

I can only make a start in replying to this challenge here, and my reply will also rely on the particular account of desert I have presented. First, the object of desert, or, in other words, what is deserved, may very well vary greatly across domains. But another dimension that varies across accounts of desert focused on the moral domain is the reason-giving force of desert. The conditional reason thesis set out earlier offers particular flexibility in this regard. The fact that someone deserves a negative response for doing something morally blameworthy does not by itself provide any reason to actually provide that response. Only together with other considerations does the fact that a person deserve something become part of even a pro tanto reason for our providing them with what they deserve. In the moral case, it might be that only if it is necessary to distribute some harm does desert become part of a reason to distribute the harm to the deserving first, for example. Thus, even in the moral case, the fact that someone deserves something negative, while having practical import, does not by itself provide a reason. If we could achieve all of the instrumental goods that typically flow from a blameworthy person’s getting what she deserves in some other way without cost, then there would be no reason at all to ensure that she gets what she deserves. If this is correct in the moral case, then it is also less difficult to accept the idea that desert, too, transposes across domains, where it also requires other background conditions to be satisfied before becoming part of a reason to provide a person with what is deserved. Thus, the commitments of accepting desert in this domain are less costly than on other accounts of desert in the moral
domain. Finally, in my view, when it comes to at least certain domains (such as the aesthetic or athletic), norms only apply in cases in which we have opted in to them, in contrast to the moral domain. It does seem problematic to think that we might find ourselves falling short of athletic standards, say, that we did not sign on for and which we might have no control over meeting, and yet nevertheless deserve a negative response. But the view does not commit us to this implausible conclusion. Of course, to accept this account, much work would have to be done in articulating the background conditions that would make desert part of a reason to provide what is deserved, along with identifying the relevant range of objects of desert. I have only sketched a direction for doing so here.

8. The Third Challenge: Too Soft or Too Splintered?

I conclude by briefly considering a final challenge, and it takes us full circle back to the beginning. After all, one of the main motivations propelling quality of will theorists is that it allows us to be more inclusive about a variety of cases, including those discussed earlier. Are we left without any sort of robust response to psychopaths, those motivated by unconscious but insidious biases, and so on? For, as we have seen, at least some quality of will theorists are in agreement with control theorists that—properly described—those individuals are not deserving of a setback to their interests. But now, if responsibility in the accountability sense, responsibility that is essentially connected to the appropriateness of the reactive attitudes comes along with desert, then they cannot be responsible in that sense, nor is it appropriate to react with reactive attitudes.

I take this concern quite seriously. It is open to say, on the approach I’ve offered, to say that bad quality of will, even in the absence of control, is worthy of more than mere aretaic judgments. It is worthy of what Watson calls “hard feelings” (2011). The distinction between resentment and hard feelings can seem too fine (and I have worried about it myself.) But I think that it is possible to make it conceptually, even if in practice these are often difficult to tease apart. These are feelings, though, that do not implicate desert. And once it is clear that such agents do not have control, I believe it is much harder to see them as being such that they can be held to account. It might be that an overly exclusive focus on the reactive attitudes not only has the potential to obscure questions of desert, but also of the very idea of whether it is appropriate to make demands of others. Rejecting the idea that such individuals are responsible for their actions in a sense that makes them accountable is perfectly compatible with protest being appropriate and even required.

One might then object that this acknowledgement ushers in an unacceptable and artificial splintering of concepts that is not reflective of a seamless and holistic practice of responsibility.

---

31 It is also the case that if the correct conception of responsibility across domains includes a substantial notion of control, then it will be more plausible to accept the idea of desert in other domains, as well.
32 See Talbert (2008), Sripada (2017), Rudy-Hiller (manuscript) for eloquent articulations of this motivation.
33 Shoemaker (2015) distinguishes between three kinds of responsibility, and while I disagree that the phenomena he has identified are really kinds of responsibility, I am sympathetic to the idea that we can make principled distinctions of this kind.
But I believe that while our practices and responses to each other are messy, there are genuine distinctions we can make that reflect distinct agential properties, and that is of great importance to make them.

9. Conclusion

In concluding, I put forward the idea that the debate about responsibility is proceeding in a more holistic way than the debate over distributive justice appears to be. When it comes to moral responsibility, we are homing in on the concepts in which we are interested as much as the conceptions. But it’s crucial to do so: It’s not that we can simply argue about whether or not psychopaths are responsible, or about what the conditions are for being blameworthy for one’s implicit attitudes, independently of making sure we are taking the same things to be at stake.

References


Fischer, John Martin; Kane, Robert; Pereboom, Derk, and Vargas, Manuel. (2007). Four Views on Free Will (Oxford University Press).


