**It’s (almost) all about desert: on the source of disagreements in responsibility studies**

Fernando Rudy-Hiller

This is a pre-peer reviewed version of the article published in *The Southern Journal of Philosophy*

https://doi.org/10.1111/sjp.12405

**Abstract**

In this paper I discuss David Shoemaker’s recently published piece “Responsibility: the state of the question. Fault lines in the foundations.” While agreeing with Shoemaker on many points, I argue for a more unified diagnosis of the seemingly intractable debates that plague (what I call) “responsibility studies”. I claim that, of the five fault lines Shoemaker identifies, the most basic one is about the role that the notion of deserved harm should play in the theory of moral responsibility. I argue that the deep divide between those theorists who affirm, and those who deny, that moral responsibility is essentially about the justification of desert thus understood can be traced to the disagreement about whether the focus on the reactive attitudes *by itself* entails that moral responsibility has nothing to do with traditional questions about desert and free will. I then show that the seeming intractableness of the other four fault lines Shoemaker identifies is expectable and explicable in light of this more basic disagreement. After this diagnostic work, I conclude by suggesting a solution to the “morass” that has taken over responsibility studies: theorists working in the field should acknowledge that it has effectively bifurcated into two discrete subareas, which I suggest calling “retribution studies” and “interpersonal studies”.

**1. Introduction**

David Shoemaker’s “Responsibility: the state of the question. Fault lines in the foundations” (2020), recently published in this journal,[[1]](#footnote-1) makes a valuable contribution in attempting to make sense of the messy theoretical field of what we can call “responsibility studies.”[[2]](#footnote-2) Shoemaker’s diagnosis is that the messiness emerges due to five different “fault lines” about certain basic assumptions having to do both with the object of study of the field and the appropriate theoretical framework that should be adopted to address it, fault lines that can be traced to Strawson’s “Freedom and resentment” (1962/2003). These fault lines concern: (1) the field’s data set (what kinds of responses count as holding someone responsible); (2) the data set’s theoretical status (whether these responses provide *evidence* of the nature of responsibility and responsible agency or rather *constitute* what responsibility and responsible agency are); (3) the question whether desert is among the appropriateness conditions of responsibility-characteristic responses; (4) the question whether these responses are essentially (or even exclusively) backward-looking rather than forward-looking; and (5) the question whether there is only one type of moral responsibility or many.

While agreeing with Shoemaker on many points, my aim in this discussion will be to introduce more unity in the diagnosis of the fault lines he identifies in the hope that, once we clearly isolate the more basic source of disagreements, the resulting disputes will become more tractable largely because it will become apparent that in many cases disagreements result from the fact that different theorists are theorizing about different things. In particular, I will argue that, of the five fault lines mentioned above, the third one—concerning the role that desert should play in the theory of responsibility—is the most basic and that the other four can be derived from it. If this is correct, the upshot is that different theorists can focus their attention on a single question (or at least a cluster of related questions) in order to find out whether they really are disagreeing or simply theorizing about different things. My final suggestion will be that the field of responsibility studies should (and most likely will) bifurcate into two discrete subareas, which we might call “retribution studies” and “interpersonal studies”.

**2. A core concept of moral responsibility**

To begin with, it’s worthwhile to emphasize that there *is* a core concept of moral responsibility that all theorists Shoemaker cite will accept and in virtue of which there is such a thing as a field meriting the label “responsibility studies”. It’s this:

**Moral responsibility.** Moral responsibility is about the evaluations and reactions that are appropriately triggered by how well or badly people’s conduct coheres with certain norms and expectations, particularly (but not exclusively)[[3]](#footnote-3) those of morality.

This definition is sufficiently broad so as to encompass all the different approaches to responsibility Shoemaker surveys in his article.[[4]](#footnote-4) For starters, “conduct” should be understood very capaciously, since it must include things like intentional actions and omissions; unintentional actions and omissions; states of awareness (what one notices and forgets); spontaneous emotional reactions; and the expression of character traits,[[5]](#footnote-5) all of which have been thought to be fitting objects of responsibility assessments. Also, the terms “evaluations” and “reactions” are so general that they can encompass things as different as judgments of moral worth; judgments that one no longer stands in relations of mutual regard with another person; aretaic appraisals; requests for justification and recognition of fault; emotional responses; attempts at moral address; interpersonal dynamics that aren’t sanctions such as withdrawals of trust and friendly attitudes; sanctions properly; and punishment. In short, the whole panoply of things that constitutes the data set occupying responsibility theorists. Thirdly, this definition makes room for investigating not only what is involved in *holding* people responsible but also what is involved in people *being* responsible, since almost everyone after Strawson accept that the evaluations and reactions in question provide the key for establishing what the distinctive capacities of responsible agents are—even if there is disagreement about whether responsible agency provides an independent standard against which we can gauge the appropriateness of these evaluations and reactions or not.

It should also be noted that the proposed definition strides over all five of Shoemaker’s fault lines, since it’s silent on exactly which evaluations and reactions are at stake; on whether these evaluations and reactions constitute what responsibility is or merely provide evidence for the same; on what their appropriateness conditions are (including whether they involve desert or not); on whether they are backward- or forward-looking; and on whether there is a single type of responsibility or several.

The importance of isolating this core common concept of moral responsibility is that in this way we defuse skepticism about the existence of a sufficiently unified enterprise that merits the label “responsibility studies” at all, which isn’t an unreasonable conclusion to draw after reading Shoemaker’s alphabet-long list of the diverse approaches to responsibility (207). With this core concept at hand, let’s investigate how the five fault lines Shoemaker identifies are generated.

**3. Desert as the basic fault line**

I will now argue that the fundamental disagreement among people working in responsibility studies is about the place (if any) that the notion of deserved harm or, alternatively, deserved suffering, ought to have in a theory of moral responsibility, which corresponds to the third fault line Shoemaker identifies. As Shoemaker characterizes it, the disagreement here is twofold (221-2): on the one hand, philosophers disagree on whether blame really has a harmful aspect of the sort that requires moral justification; on the other, they disagree about what moral justification can be offered for those instances of blame that are harmful (assuming there are any). I characterize the dispute more succinctly in terms of the fundamental divide between those who accept and those we deny the following claim:

**Desert.** Moral responsibility has something intrinsically to do with the justification of the harm or suffering deserved by those who commit wrongdoing.

The thought here is simple enough: if you did something wrong or bad, now you deserve (in some sense) to suffer (in some way) in response. This is the essence of retributivism.[[6]](#footnote-6) It’s true that retributivists disagree on many points (Nelkin 2016), for instance on whether the fact of wrongdoing makes it permissible or obligatory to inflict harm or simply warrants a contrastive reason, i.e., if you need to harm anyone, harm the blameworthy rather than the blameless (the “deontological question”). They also disagree on whether it’s intrinsically good that the blameworthy suffer and on how to characterize the goodness in question (“the axiological question”). In addition, retributivists disagree massively on which kinds of harms or sufferings are at stake, going from the pain of guilt (Rosen 2015; Carlsson 2017) all the way to eternal damnation in hell (Strawson 1994). And, of course, as Shoemaker points out, they disagree on how the very notion of desert should be understood (222). And yet, accepting or rejecting **Desert** is *the* fundamental theoretical decision-point for those working in responsibility studies. If you think something like **Desert** is right, this will color your choices at all the other theoretical decision-points Shoemaker describes—and the same is true of you reject it. Going forward, I will call those philosophers who accept **Desert** “retributivists” and, for reasons that will become apparent in a moment, I will call those who reject it “Strawsonians.”

How was the divide between retributivists and Strawsonians generated? As with almost everything else in this domain, it goes back to “Freedom and resentment.” As Shoemaker nicely puts it, Strawson’s chief aim in that paper was to “free us from the quest for freedom” (212), that is, to convince philosophers that questions about moral responsibility can and should be divorced from questions about free will and determinism. To accomplish this, Strawson focused on the kind of reactions that are the staple mark of engaged interpersonal relationships—the reactive attitudes. The fundamental divide on the post-Strawsonian era is, I will argue now, a divide over whether this focus on the reactive attitudes *by itself* entails that moral responsibility has nothing to do with desert as traditionally understood and, given the central place that desert plays in the free will debate, that moral responsibility and free will are two distinct and possible unrelated topics.

On the one hand, many people—those I have called retributivists—tend to conceive the reactive attitudes as an interpersonal analogue of institutionalized sanctions and punishment and therefore assume we can continue asking the same questions that traditionally have been asked in the free will debate by appealing to them (e.g., Wolf 1990: 20-1; Wallace 1994; Fischer and Ravizza 1998: 5-8; Watson 2004: 278; Pereboom 2014: 2; Nelkin 2016). On the other hand, many people—those I have called Strawsonians—think that, once we focus on the reactive attitudes, *we* *thereby* *have changed the topic*. The crucial point for this second group of theorists is that the reactive attitudes are *categorically different* from sanctions and punishment and therefore their appropriateness conditions—including the agential capacities required for being aptly targeted by them—are categorically different as well (Hieronymi 2004, 2019; Scanlon 2013a, 2015; Shoemaker 2013; Talbert 2016; Smith 2019).[[7]](#footnote-7)

Let me elaborate on the reasons why Strawsonians think this is the case. First, some have noted that many of the paradigmatic reactive attitudes, including resentment and indignation, are *non-*voluntary in the sense that one can’t decide to adopt them at will, much in the same way in which one can’t believe at will. Rather, one simply *finds* oneself resenting or being indignant at someone after judging that they have shown disregard or ill will to oneself or another, just as one finds oneself believing that *p* after judging the *p* is the case. Thus, the reactive attitudes are fundamentally unlike sanctions and punishment, which are voluntary activities and, as such, are subject to the same standards of justification that apply to any action that intentionally burdens others (Hieronymi 2004, 2019).

Second, others have argued that, even if it’s true that the reactive attitudes can be burdensome to their targets, they aren’t a species of sanctions or punishment because they aren’t adopted *in order to* make their targets suffer “but for reasons having to do with our own concern with our relationships with them” (Scanlon 2015: 93). In other words, the primary function of the reactive attitudes is “interpersonal calibration”, that is, they are grounded on our concern with having attitudes toward people that are suitably attuned to the treatment they dispense us and others we vicariously identify with.[[8]](#footnote-8)

Third, even if, as some people have claimed, blaming reactive attitudes often incorporate thoughts and desires that their targets suffer (McGeer 2013; Rosen 2015), this entails neither that they are sanctions nor that they necessarily involve a disposition to engage in sanctioning behavior (Wallace 1994; Watson 2004). For instance, even if upon learning that a sadistic crime has been committed against a particularly vulnerable person (a child, say) I find myself experiencing not only intense indignation but also a strong desire that the perpetrator suffer, this doesn’t entail that my reactions are sanctions nor does it entail that I would, even if the circumstances were propitious, act on that desire or even think that doing so would be appropriate. Therefore, the preparedness to acquiesce in the infliction of suffering (to use Strawson’s phrase) doesn’t incur the justificatory burdens of actually inflicting suffering. Moreover, even if some people—plausibly, all of us at some point or another—do employ blaming responses with the aim of causing pain, this doesn’t show that we need a moral justification for the reactive attitudes of the sort that is needed for sanctions and punishment. As Shoemaker aptly puts it, “People can be fairly mean in their retribution. But that is not what angry blame itself wants; it is instead what angry *people* sometimes want” (224). That is, the presence of retributive impulses in people doesn’t show that the reactive attitudes are by nature vehicles for the expression of those impulses; therefore, they don’t incur the justificatory burdens of retributive practices.

Fourth and finally, given that accountability relations among adults are premised on their status as equal members of the moral community, they possess *mutual* authority to make demands on one another (Darwall 2006). At the same time, their equal status entails that they lack authority *over* one another to inflict harm in response to disappointed demands and expectations. So while I can certainly respond to your wrongdoing in ways you have reason to dislike (like giving the cold shoulder to an untrusty friend), these responses don’t have the status of sanctions because both formal and informal sanctions (e.g., legal punishment administered by the state, a parent grounding their children) presuppose a form of authority over another that is simply absent in moral relations among equals (Shoemaker 2013).

I think this set of reasons constitutes a fairly weighty case in favor of the Strawsonian position. However, we should be very careful in stating exactly what follows from it. It does *not* follow that moral responsibility is categorically different from questions about sanctions, punishment and desert, full stop. Rather, it only follows that, *if* moral responsibility is understood in terms of the reactive attitudes, *then* moral responsibility is categorically different from questions about sanctions, punishment and desert (where desert is understood as whatever justifies harm or suffering in response to wrongdoing). And yet this is a very substantive conclusion, since it serves both to largely explain the mess that many people feel has taken over responsibility studies and to show how to clean it up. It explains the mess because, in failing to recognize that this fundamental divide between retributivists and Strawsonians actually entails a radical change of topic, it has become very difficult (if not impossible) for philosophers to evaluate the mutually contradictory claims that are made about the conditions for moral responsibility (Zimmerman 2015; see section 4 below). And it contributes to cleaning up the mess because it suggests the following advice: if you are interested in the interpersonal reactions and dynamics that characterize moral relations and, more broadly, engaged interpersonal relationships in general, then you should stop worrying about issues of free will, control,[[9]](#footnote-9) and desert. If, on the other hand, you are interested in free will, control and desert, i.e., on the agential condition that could justify the infliction, or simply the presence, of harm or suffering in response to wrongdoing, then you should *not* focus on the reactive attitudes as your preferred gloss on what moral responsibility is. How, then, should you understand moral responsibility if you are interested in these things? My suggestion: focus on either divine or legal punishment.[[10]](#footnote-10)

In his final assessment of the dispute between those I have called retributivists and Strawsonians, Shoemaker writes: “These are indeed radically different approaches, and it is unclear how they may ultimately be brought together” (225). I agree with Shoemaker that these are radically different approaches, but the crucial question is not whether they can be brought together but rather whether defenders of each can be brought to see that they are actually theorizing about different things. In other words, these approaches are radically different because they have different objects of study. Strawsonians are concerned with explaining the relevance of the reactive attitudes for social beings like us for whom the regard of others inescapably matters (Hieronymi 2020) and investigating those attitudes’ “normative felicity conditions” (Darwall 2006: 4-5, quoted by Shoemaker, p. 224), whereas retributivists are interested in what kind of agents we must be for “truly” deserving harm or suffering in response to wrongdoing (Pereboom 2014) or for deserving the “special condemnatory force” of blame and its associated harms (Wolf 1990, 2011; Wallace 1994, 2011; McKenna 2019).

I am confident that substantial agreement can be reached on this point. That is, I think that people can be led to agree that the reactive attitudes are just too different from sanctions and punishment and therefore have categorically different appropriateness conditions, and thus they can be led to agree that if moral responsibility is understood in terms of the former then it doesn’t have anything to do with desert and free will/control.[[11]](#footnote-11) Before exploring in the final section the consequences this imagined agreement would have for the future of responsibility studies, I will explain in the next one how the other four fault lines Shoemaker identifies can be derived from this basic divide between retributivists and Strawsonians.

**4. The other fault lines**

I will now argue that, once this fundamental divide between retributivists and Strawsonians is in place, the other four major fault lines Shoemaker identifies are fully expectable and explicable.[[12]](#footnote-12) I don’t mean the following to be an accurate reconstruction of how these other fault lines were actually generated; rather, my point will be that what gives these other disputes their air of intractableness is the foundational disagreement about the role that desert should play in the theory of responsibility. That is, these disputes *seem* to embody intractable disagreements *because* what drives them is a basic unclarity on the part of theorists about what their object of study is supposed to be. By contrast, once we are clear about this—once we distinguish clearly between retributivism and Strawsonianism—these disputes will lose much of their apparent recalcitrancy (though not necessarily their interest).

*Composition of the data set*

Shoemaker correctly argues that there is a major dispute “over the nature of the data set and how exactly to determine its contents,” that is, a dispute about “[w]hat counts as the relevant way of ‘holding responsible’” (217). This is exactly what we should expect in light of the underlying disagreement between retributivists and Strawsonians.[[13]](#footnote-13) That is, given a fundamental disagreement about whether responsibility theory is essentially about explaining and justifying how people can come to deserve harsh treatment—if only under the guise of the “opprobrium” of blame (Wallace 2011)—then we should expect that apparently intractable disputes will emerge downstream about whether the reactions that matter for theorizing about responsibility are those that necessarily involve (or presuppose or license) a harmful aspect or not.[[14]](#footnote-14)

If you fall in the retributivist side, you will give an affirmative answer to this question. This is why people with retributivist leanings focus on a “harsh” construal of blame (Wolf 2011; Wallace 2011; McGeer 2013; Rosen 2015) and, more specifically, on resentment, because they assume that resentment is, by its nature, harmful and the interpersonal analogue of sanctions and punishment—what gives blaming reactions is characteristic force or depth (Wolf 1990: 20-1; Rosen 2015). This is also why they focus on sanctioning activities (Wallace 1994; Watson 2004) and assume that there is a continuum between reactive attitudes, sanctions, and punishment (Fischer and Ravizza 1993: 19; Wallace 1994: 51; Vargas 2020: 413) and between moral and legal responsibility (Brink and Nelkin 2013). By contrast, if you are a Strawsonian you will deny all of the foregoing and will be progressively drawn to a position that welcomes into the theory pretty much *all* the responses that are relevant for managing our interpersonal relationships irrespectively of whether they can be categorized as species of praise and blame (Tognazzini 2015: 38-9; Watson 2019: 224), excluding only those that *indisputably* are sanctions and punishment—most clearly, negative consequences imposed by some person or body in a position of authority in response to the violation of a norm (Scanlon 2013b; Hieronymi 2019).

The relevant point for present purposes is that in all likelihood the dispute about the data set will become far less intractable once different theorists get clear about whether they are interested in desert or not, since it will then be much clearer what responses are the ones they should focus on.[[15]](#footnote-15) This doesn’t mean that all disputes in this regard will come to an end; on the contrary, interesting *internal* debates will likely emerge within the retributivist and Strawsonian camps. A case in point is Shoemaker’s contention that if we include hurt feelings in the date set, Strawsonians may be forced to abandon their fixation with quality of will, i.e., the view that all responsibility-relevant reactions are sensitive to the agent’s moral regard or lack thereof (213-4). This is an interesting suggestion worth addressing, but if we start from a basic agreement about what we are supposed to be theorizing about—e.g., the spontaneous responses that arise when interpersonal expectations are unmet—it should be easier to ascertain whether the response in question (e.g., hurt feelings) should be included in the data set or not.[[16]](#footnote-16)

*Response-dependent vs. response-independent theories*

Another major dispute according to Shoemaker concerns the theoretical status of the data set: do responsibility-relevant reactions play a *constitutive* or an *evidential* role regarding moral responsibility? (218) In other words, is the property of being responsible for something a response-dependent or a response-independent property? Once again, my suggestion is that the apparent intractableness of this debate owes a lot to an upstream disagreement about the role that desert should play in the theory of responsibility. If you are a retributivist, you will think that a response-independent theory it’s the only viable option, because it’s crucial for you to find out whether people really have the requisite agential capacities (e.g., control, free will) for truly deserving harm in response to wrongdoing. A response-dependent theory of desert makes no sense, since only those agents who are *actually* culpable truly deserve harm or suffering (on the retributivist’s view). By contrast, if you are a Strawsonian the decision to adopt either methodology isn’t settled as a matter of course; both may appear attractive for different reasons. A response-independent theory is attractive because it allows us to preserve the view that we (not only as individuals, but as community of appraisers) can make genuine mistakes in our attributions of responsibility, for example mistakenly thinking that certain agents express morally relevant ill will when they are actually incapable of such a thing (psychopaths, perhaps) (Fischer and Ravizza 1993: 18). On the other hand, a response-dependent theory is attractive for exactly the reason Shoemaker (220) gives: since (at least most of) responsibility-characteristic responses are triggered by whatever us human beings regard as relevant for evaluating each other’s quality of will, it’s plausible to think that what it is to be responsible for something is simply to be targeted by one of these responses under appropriate circumstances. These are genuine and interesting theoretical disputes within the Strawsonian camp, but in the end the distinction between a response-dependent and a response-independent approach won’t be as important as it is for retributivists. Once you agree on the basics, i.e., that your object of study is the interpersonal reactions that characterize engaged human relationships, it will be less consequential if different theorists disagree on the exact theoretical status of these reactions.[[17]](#footnote-17) By contrast, if we haven’t agreed on the basics—if we are discussing *across* the fault line separating retributivists and Strawsonians—the disagreement between response-independent and response-dependent views will naturally seem not only extremely important but intractable. The essential point for current purposes is again that the apparent intractableness of this disagreement is really an artifact of the previous and more fundamental dispute about the role of desert in the theory of responsibility*.*

*Backward-looking vs. forward-looking views*

I will suggest a very similar diagnosis of the fourth fault line Shoemaker identifies. If one is a Strawsonian, the need to establish a sharp distinction between the backward-looking and the forward-looking aspects of responsibility should be less urgent than if one is a retributivist, since ordinary interpersonal responses are an inextricable mixture of both. So while these responses are certainly triggered by what their targets did (or simply by what they are) and, on the side of blaming responses, they surely have as one of their central functions that of registering that an offense was suffered by oneself or another, they have all sorts of other functions that don’t seem to be exclusively backward-looking in nature. These include protesting wrongdoing (Smith 2013), bolstering the victim’s status as someone who merits consideration (Talbert 2012), signaling commitment to norms (Shoemaker and Vargas 2019), initiating a moral conversation (McKenna 2012), scaffolding people’s moral sensitivities (Vargas 2013; McGeer 2019), recalibrating one’s relationships with others (Scanlon 2008), etc. Thus, for Strawsonians the distinction between the backward-looking and the forward-looking aspects of responsibility-characteristic reactions will be blurry and mostly unimportant.[[18]](#footnote-18) On the other hand, for retributivists it is crucial to emphasize that the deservingness conditions of these reactions are exclusively backward-looking, since it makes no sense to say that someone deserves harm or suffering in the “basic sense” on account of something that lies in the future (Pereboom 2014), just as it makes no sense to say that an athlete deserves a medal because it will bolster her athletic capabilities (Doris 2015).[[19]](#footnote-19) Once again, my suggestion is that the apparent urgency of settling this debate is really an artifact of a failure to clearly distinguish the very different objects of study with which retributivists and Strawsonians are concerned.

*Monism or pluralism about responsibility*

We arrive finally at the one place in responsibility studies where the word “morass” is most aptly applied: the dispute about the “faces” or “types” of responsibility (Shoemaker’s fifth fault line). Once more, I contend that the seeming intractableness of this debate owes a great deal to a confusion about the very different theoretical concerns of retributivists and Strawsonians. If you are a retributivist, you will naturally think either that there is only one legitimate type of responsibility (Levy 2005) or that, while there might be several, the truly important one is the one linked to deserved harm or suffering (Levy 2011: 3; Pereboom 2014: 2; Nelkin 2016). Moreover, you will insist that the relevant type of responsibility is the one that connects to, or is at issue in, the free will debate (Levy 2011; McKenna 2012; Pereboom 2014; Nelkin 2016). Since Watson (2004) introduced the distinction between two faces of responsibility, retributivists have taken it as a given that they are concerned with the *accountability* face, since it has been assumed (again, since Watson) that accountability is necessarily linked to deserved harm (see Nelkin 2016 for a very clear articulation of this view). But here come the confusion: as I noted above, retributivists often describe accountability *in terms of the appropriateness of the reactive attitudes* (Wallace 1994; Fischer and Ravizza 1998: 5-8; Nelkin 2016), assuming that the latter embody their central concern about deserved harm. This sowsconfusion because, obviously, Strawsonians insist that they are *also* concerned with the appropriateness conditions of the reactive attitudes but deny pretty much all that retributivists assert about what these conditions involve (Scanlon 2013a; Talbert 2016). Given this set up, retributivists often dismiss Strawsonians accounts on the grounds that they fail to engage with the traditional problem of free will (Levy 2011: 209; Pereboom 2014: 98-9) and, by contrast, Strawsonians accuse retributivists of misunderstanding the nature of these attitudes and, by extension, of responsibility itself (Scanlon 2013b: 108; Hieronymi 2019).

Thus, I disagree with Shoemaker that the fault line here is generated by an ambiguity in the notion of “quality of will” (228). To be clear, what I deny is that the apparent intractableness of the debate about the types of responsibility is due to a confusion about the different ways in which “quality of will” can be interpreted. I claim, rather, that it’s due to a confusion about the different objects of study of retributivists and Strawsonians. This doesn’t deny that there are different ways of construing what quality of will involves nor does it deny that they generate important disagreements *among* Strawsonians, but it does entail that the source of the seeming intractableness of the debate about the faces of responsibility lies elsewhere.

If you are a Strawsonian you should be more open to plurality about responsibility, especially if you accept that your object of study involves an enormous swath (maybe all!) of the responses that characterize engaged human relationships. It would be astonishing if all these responses shared the same set of appropriateness conditions, including the same set of agential conditions. Interesting debates will remain about whether there is one organizing set of conditions or not (Smith 2012, 2015; Shoemaker 2015), but this shouldn’t make Strawsonians go to the barricades, as they will have to if they continue to be engaged with retributivists in the debate about what the “true” or “most important” type of responsibility is.[[20]](#footnote-20)

Notice, finally, that the distinction between retributivists and Strawsonians isn’t the same as the distinction between accountability and other types of responsibility. As many authors have noted, the essence of accountability consists in being accountable *to* someone, that is, it consists in being subject to other people’s legitimate demands and expectations and to the reactions that are appropriate when these demands and expectations are disappointed or exceeded (e.g., Watson 2004, 2019; Darwall 2006; Nelkin 2016). This presupposes “the capacity for moral reciprocity or mutual recognition that is necessary for intelligibly holding someone accountable to basic moral demands and expectations” (Nelkin 2015: 363). Crucially, however, being accountable to someone in this sense isn’t the same as being subject to another’s authority to administer sanctions and punishment (Shoemaker 2013), nor does it entail anything about the goodness of the blameworthy suffering some harm or about the obligatoriness, permissibility, or just preferableness that they do (Scanlon 2013b). Therefore, Strawsonians have every right to a respectable notion of accountability and thus the divide between retributivists and Strawsonians can’t be equated to the dispute about accountability per se.[[21]](#footnote-21)

**5. The future of responsibility studies**

To sum up, I have argued that the deep theoretical and methodological divisions that characterize the field of responsibility studies can all be traced to a foundational disagreement about the role that desert (understood as deserved harm or suffering) should play in theorizing about moral responsibility. Strawsonians, I have argued, have a very strong case that once we focus on the ordinary reactions that characterize engaged human relationships—the reactive attitudes—we thereby have changed the topic vis-à-vis the traditional free will problem and its associated conception of moral responsibility. Failure to clearly recognize this point by both retributivists *and* Strawsonians has led to seemingly intractable disputes down the road. The solution I propose is thus simple: theorists working on responsibility studies should acknowledge that the discipline has effectively bifurcated into two discrete subareas, which I suggest calling “retribution studies” and “interpersonal studies”.[[22]](#footnote-22)

These are subareas of a single philosophical field—what I have called “responsibility studies”—because they agree on the basic definition of moral responsibility I gave in section 2. But they are discrete subareas nonetheless because they are concerned with categorically different reactions to morally loaded conduct with categorically different appropriateness (including agential) conditions. Retribution studies will retain the tight link between moral responsibility and the questions that have informed the traditional free will debate, and thus will continue focusing on understanding the (im)possibility of persons coming to deserve good or bad things simply on account of what they have done. Theorists working in this subarea will benefit from clearly acknowledging—as some philosophers have already done[[23]](#footnote-23)—that the conception of moral responsibility they are concerned with is closely connected to the one at stake in discussions of criminal responsibility and in theistic debates about free will, rather than to the one pioneered in “Freedom and resentment”. On the other hand, interpersonal studies will be in charge of mapping out, understanding, and laying out the appropriateness conditions of pretty much *all* responses that characterize engaged human relationships (Tognazzini 2015; Watson 2019). A beneficial consequence of this extremely wide focus is that it will naturally admit into our theorizing people who are usually left aside in responsibility studies—those Shoemaker calls “marginal agents”—since “normal adults” aren’t the only people who interact with and react to each other (231-2).[[24]](#footnote-24)

My diagnosis and proposed solution to the chaos that has taken over responsibility studies isn’t intended as a magic wand that will end all disputes in one fell swoop. My hope is, rather, that once there is widespread acknowledgement that the fundamental division between retributivists and Strawsonians amounts to a cleavage between two separate subareas of study, the remaining disagreements *within* these subareas will become more focused and less intractable (though not less interesting), since there will be internal agreement on the basic foundational issue, i.e., what the object of study is. Thus, rather than something to be lamented, this partition of the field could produce a profitable division of labor between the two subareas sketched above. So I think Shoemaker’s metaphor of “fault lines” is ultimately misleading, because it suggests a bad sort of division (no one would characterize a well-functioning division of labor as being “riven by fault lines” [205]). It’s bad when something that should be unified is fractured; but it’s not bad (it’s actually good!) when something that is basically diverse is recognized as such.

It remains to be seen whether this proposed solution proves too optimistic. For my own part, I refuse to accept that those of us working on responsibility studies are condemned to confronting one dialectical stalemate after another.

**References**

Arpaly, N. and T. Schroeder (2014) *In Praise of Desire*. Oxford: OUP.

Brink, D. and D. Nelkin (2013) “Fairness and the Architecture of Responsibility.” *Oxford*

*Studies in Agency and Responsibility*, Volume 1. Oxford:OUP, pp. 284–314.

Burroughs, M. (2020) “Navigating the Penumbra: Children and Moral Responsibility.”

*Southern Journal of Philosophy* 58 (1): 77-101.

Carlsson, A. (2017) “Blameworthiness as Deserved Guilt.” *The Journal of Ethics* 21: 89-

115.

Clarke, R. (2016) “Moral Responsibility, Guilt, and Retributivism.” *The Journal of Ethics*

20: 121–137.

Darwall, S. (2006) *The Second-Person Standpoint*. Cambridge, MA: Harvard University

Press.

Doris, J. (2015) “Doing Without (Arguing about) Desert,” *Philosophical Studies* 172:

2625–2634.

Fischer, J. (2007) “Compatibilism.” In J. M. Fischer, R. Kane, D. Pereboom, and M.

Vargas (eds.) *Four Views on Free Will*, pp. 44–84. Oxford: Blackwell Publishers.

Fischer J. and M. Ravizza (1993) “Introduction.” In J. Fischer and M. Ravizza (eds.)

*Perspectives on Moral Responsibility*, pp. 1–41. Ithaca, NY: Cornell University Press.

Fischer, J. and M. Ravizza (1998) *Responsibility and Control: A Theory of Moral*

*Responsibility*. Cambridge: CUP.

Hieronymi, P. (2004) “The Force and Fairness of Blame,” *Philosophical Perspectives* 18:

115-148.

Hieronymi, P. (2019) “I’ll Bet You Think This Blame is About You.” *Oxford Studies in*

*Agency and Responsibility* 5: 60–87.

Hieronymi, P. (2020) *Freedom, Resentment, and the Metaphysics of Morals*. Princeton:

Princeton University Press.

Levy, N. (2005) “The Good, the Bad, and the Blameworthy.” *Journal of Ethics and Social*

*Philosophy* 1 (2): 1-16.

Levy, N. (2011) *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*.

Oxford: OUP.

McGeer, V. (2013) “Civilizing Blame.” In D. Coates and N. Tognazzini (eds.) *Blame: Its*

*Nature and Norms*, pp. 162-188, Oxford: OUP.

McGeer, V. (2019) “Scaffolding Agency: A Proleptic Account of the Reactive Attitudes.”

*European Journal of Philosophy* 27 (2): 301-323.

McKenna, M. (2012) *Conversation and Responsibility*. Oxford: OUP.

McKenna, M. (2019) “Basically Deserved Blame and its Value.” *Journal of Ethics and*

*Social Philosophy* 15 (3): 1-28.

Nelkin, D. (2015) “Psychopaths, Incorrigible Racists, and the Faces of Responsibility.”

*Ethics* 125 (2): 357-390.

Nelkin, D. (2016) “Accountability and Desert.” *The Journal of Ethics* 20: 173-189.

Pereboom, D. (2014) *Free Will, Agency, and Meaning in Life*. Oxford: OUP.

Rosen, G. (2015) “The Alethic Conception of Responsibility.” In R. Clarke, M.

McKenna, and A. Smith (eds.) *The Nature of Moral Responsibility*, pp. 65-88. Oxford: OUP.

Scanlon, T. M. (2008) *Moral Dimensions*.Cambridge, MA: Harvard University Press.

Scanlon, T. M. (2013a) “Interpreting Blame.” In D. Coates and N. Tognazzini (eds.)

*Blame: Its Nature and Norms*, pp. 84-99. Oxford: OUP.

Scanlon, T. M. (2013b) “Giving Desert its Due.” *Philosophical Explorations* 16 (2): 101-

116.

Scanlon, T. M. (2015) “Forms and Conditions of Responsibility.” In R. Clarke, M.

McKenna, and A. Smith (eds.) *The Nature of Moral Responsibility*. Oxford: OUP, pp. 89-111.

Shapiro, T. (1999) “What is a Child?” *Ethics* 109: 715–738

Shoemaker, D. (2013) “Blame and Punishment.” In D. Coates and N. Tognazzini (eds.)

*Blame: Its Nature and Norms*, pp. 100-118. Oxford: OUP

Shoemaker, D. (2015) *Responsibility from the Margins*. Oxford: OUP.

Shoemaker, D. (2020) “Responsibility: the State of the Question. Fault Lines in the

Foundations.” *The* *Southern Journal of Philosophy* 58 (2): 205-237.

Shoemaker, D. and M. Vargas (2019) “Moral Torch Fishing: A Signalling Theory of

Blame.” *Noûs*, early online <https://doi.org/10.1111/nous.12316>

Smith, A. (2012) “Attributability, Answerability, and Accountability: In Defense of a

Unified Account.” *Ethics* 122: 575–589.

Smith, A.  (2013) “Moral Blame and Moral Protest.” In D. Coates and N. Tognazzini (eds.)

*Blame: Its Nature and Norms*, pp. 27-48. Oxford: OUP

Smith, A. (2015) “Responsibility as Answerability.” *Inquiry* 58: 99–126.

Smith, A. (2019) “Who’s Afraid of a Little Resentment?” *Oxford Studies in Agency and*

*Responsibility* 6: 85-111.

Strawson, P. (1962/2003). “Freedom and Resentment.” In G. Watson (ed.) *Free Will*.

Oxford: OUP, pp. 72-93.

Strawson, G. (1994) “The Impossibility of Moral Responsibility.” *Philosophical Studies*

75: 5-24.

Talbert M. (2012) “Moral Competence, Moral Blame, and Protest.” *Journal of Ethics* 16

(1): 89-109.

Talbert, M. (2016) *Moral Responsibility*. Cambridge: Polity Press.

Tognazzini, N. (2015) “The Strains of Involvement.” In R. Clarke, M. McKenna, and A.

Smith (eds.) *The Nature of Moral Responsibility*, pp. 19-44. Oxford: OUP.

Vargas, M. (2013a) *Building Better Beings*. New York: OUP.

Vargas, M. (2020) “Negligence and Social Self-Governance.” In A. Mele (ed.) *Surrounding*

*Self-Control,* pp. 400-420. Oxford: OUP.

Wallace, J. (1994) *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard

University Press.

Wallace, J. (2011). “Dispassionate Opprobrium: On Blame and the Reactive Sentiments.”

In J. Wallace, R. Kumar, and S. Freeman (eds.) *Reasons and Recognition*, pp. 348-372. Oxford: OUP.

Watson, G. (2004) “Two Faces of Responsibility.” In *Agency and Answerability*. Oxford:

OUP.

Watson, G. (2019) “Second Thoughts.” In *Oxford Studies in Agency and Responsibility* 5:

214-239.

Wolf, S. (1990) *Freedom within Reason*. Oxford: OUP.

Wolf, S. (2011) “Blame, Italian Style.” In R. J. Wallace, R. Kumar, and S. Freeman (eds.)

*Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. Oxford: OUP,

pp. 332-347.

Zimmerman, M. (2015) “Varieties of Moral Responsibility.” In R. Clarke, M. McKenna,

and A. Smith (eds.) *The Nature of Moral Responsibility*. Oxford: OUP, pp. 45-64.

1. Subsequent references to this text will only give page numbers. [↑](#footnote-ref-1)
2. Like Shoemaker, Zimmerman (2015) also characterizes the field as a mess. [↑](#footnote-ref-2)
3. Other relevant norms and expectations include those of friendship, collegiality, marriage, etc., assuming that they can’t be neatly subsumed under the norms and expectations of morality. This suggests that we will likely face an explosion of the kinds of norms and expectations that are relevant for moral responsibility—which may, at the limit, become coextensive with the norms and expectations that regulate *all* kinds of interpersonal relations. This will be a natural result from the research agenda initiated by Strawson in “Freedom and resentment.” I will return to this point below. Notice also that “morality” here isn’t restricted to determinations of right and wrong, but include as well the notions of good, bad, virtue, and vice. [↑](#footnote-ref-3)
4. This definition doesn’t presume a “merited-consequences” conception of responsibility (Hieronymi 2019). Reactions being appropriate isn’t the same as consequences being merited. See section 3 below. [↑](#footnote-ref-4)
5. Even if one thinks that responsibility-characteristic evaluations can concern appraisals of character traits themselves (“aretaic appraisals”), one must concede that one appraises them by focusing on their outward expression in conduct (how else could you know that someone is greedy or courageous except by witnessing her perform greedy or courageous actions?). [↑](#footnote-ref-5)
6. Fischer (2007: 81-2) claims that the ordinary concept of moral responsibility is committed to retributivism. [↑](#footnote-ref-6)
7. It’s worth noting that, according to the characterization I have offered, Strawson himself was not a Strawsonian, since he explicitly held the view that the reactive attitudes are *not* categorically different from punishment. He wrote: “So the preparedness to acquiesce in that infliction of suffering on the offender which is an essential part of punishment *is all of a piece* with this whole range of attitudes of which I have been speaking” (Strawson 1962/2003: 90, italics added). As it sometimes occurs, I think this is a case of a philosopher not fully understanding the consequences of his own view. [↑](#footnote-ref-7)
8. This applies to interpersonal dynamics that *give expression* to non-voluntary reactive attitudes, such as giving someone the cold shoulder or refusing to help an uncooperative person with their projects. [↑](#footnote-ref-8)
9. I admit that it’s more controversial to claim that once we adopt the Strawsonian position not only free will but also control becomes dispensable. In work in progress I argue in detail that control understood in terms of reasons-responsiveness isn’t a condition of moral responsibility when the latter is conceived along Strawsonian lines. [↑](#footnote-ref-9)
10. Galen Strawson (1994) explicitly links moral responsibility to divine punishment, while Zimmerman (2015: 58) explicitly links the kind of moral responsibility that requires control to legal punishment. See also Watson’s (2004: 280) remark that moral accountability can be conceived as a “legal-like practice.” Brink and Nelkin (2013) sketch a theory of moral responsibility based on the presumption that moral and criminal responsibility share the same agential conditions and differ only on the norms and responses that characterize each. Although they mistakenly assume that moral responsibility so conceived can be characterized in terms of the reactive attitudes, they are on the right track in noticing the tight parallels between a retributivist conception of moral responsibility and a retributivist conception of criminal responsibility. [↑](#footnote-ref-10)
11. Theorists like Clarke (2016), Carlsson (2017), and McKenna (2019), who think that desert is relevant for the reactive attitudes not because the latter are a species of sanctions and punishment but because they actually cause suffering to their targets (e.g., the suffering of not being able to interact with others unimpededly or the suffering of guilt) regardless of the intentions of those who deploy them, will vigorously disagree with this conclusion. In work in progress I defend the claim that the fact that the reactive attitudes cause suffering isn’t enough for securing the relevance of free will/control for blame attributions. Following Scanlon (2015), I argue there that people don’t have an unconditional right to expect good will from others; rather, they can demand good will on the condition that they show it themselves. When they don’t, they lack grounds for complaining about the withdrawal of good will the reactive attitudes entail (which includes the suffering such withdrawal often causes to their targets), irrespectively of whether the wrongdoing in question was freely willed or not. [↑](#footnote-ref-11)
12. I accept for the sake of argument that these fault lines are indeed the most important extant disputes in the field of responsibility studies. [↑](#footnote-ref-12)
13. Presumably Shoemaker will agree on this point, since one of his glosses of this fault line is in terms of the question “what are we supposed to be theorizing about in the first place?” (217). [↑](#footnote-ref-13)
14. For a related diagnosis, see Vargas (2020: 413). [↑](#footnote-ref-14)
15. Don’t fall into the trap of thinking that the distinction between retributivists and Strawsonians is simply another name for the dispute about the “faces” of responsibility. I elaborate on this point below when discussing Shoemaker’s fifth fault line. [↑](#footnote-ref-15)
16. Just for the record, I think Shoemaker’s (214) example of someone whose feelings are hurt because someone she considered a friend threw away a gift of hers can be squarely understood as a case in which the second person evinced lack of regard—understood, precisely, as a lack of concern for the other person’s feelings. The first person could confront the second by asking, “how would you like if someone threw away a gift of yours, eh?” [↑](#footnote-ref-16)
17. I thus think, contra Shoemaker, that *both* response-dependent and response-independent Strawsonians will have to engage in “down and dirty empirical work” (220) to figure out the nature of the responses they claim to be investigating. My point, again, is that once you agree on the basics many of the theoretical disputes Shoemaker describes will be less divisive than he thinks. [↑](#footnote-ref-17)
18. I say “mostly” because Strawsonians will surely agree on Strawson’s point that the reactive attitudes aren’t (exclusively) tools for social control. [↑](#footnote-ref-18)
19. This isn’t to deny that interesting intramural disputes will remain in the retributivist camp about whether the basic commitments of retributivism can be reconciled with some forward-looking elements, such as the improvement of the wrongdoer through blaming and sanctioning activities (Vargas 2013; Pereboom 2014; McGeer 2019). However, the presence of forward-looking functions such as this doesn’t impugn the basic commitment of retributivism, viz., that the desert basis of responsibility-characteristic reactions only admits backward-looking elements (Pereboom 2014). Thanks to an anonymous referee for pressing for clarification. [↑](#footnote-ref-19)
20. Though see Shoemaker’s important observation that recognizing the plurality of responsibility is essential for respecting “marginal agents” (232). [↑](#footnote-ref-20)
21. It’s worth noting that some Strawsonians mischaracterize their own position as being about a form of responsibility basically different from accountability (e.g., Arpaly and Schroeder 2014: Ch. 7). [↑](#footnote-ref-21)
22. These are fields of study that fall squarely within ethics (or within the intersection of ethics and metaphysics, in the case of retribution studies), so I am puzzled by Shoemaker’s suggestion that responsibility theory may “become its own substantial and distinctive area of philosophical inquiry along the lines of other well-established areas of philosophy like ethics, metaphysics, and epistemology” (233). [↑](#footnote-ref-22)
23. See footnote 10 above. [↑](#footnote-ref-23)
24. This will also open the door to a more sustained study of the sorts of reactions that are appropriate regarding children’s conduct. See for instance Shapiro (1999) and Burroughs (2020). [↑](#footnote-ref-24)