Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame

David Shoemaker
Tulane University

There are many contested features of P. F. Strawson's "Freedom and Resentment." But the feature most people reject is Strawson's response-dependent view of moral responsibility.[1] By leaning on our interpersonal responsibility responses (reactive attitudes such as resentment, indignation, and guilt) and what tends to suspend them in order to account for "*all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice" (Strawson 2003, 91; emphasis in original), Strawson maintains that being responsible is a

I presented earlier versions of this essay at a Gothenburg Responsibility Project workshop and at the University of Arizona's Freedom Center. I am grateful to the insightful feedback provided by audience members at both presentations, including Dana Nelkin, Manuel Vargas, Paul Russell, Sofia Jeppsson, Per Milam, Carolina Sartorio, Connie Rosati, Steve Wall, David Schmidtz, Don Fallis, Dan Shahar, and Robert Wallace. I received excellent comments on earlier drafts of the essay from Michael McKenna, Chandra Sripada, Nathan Stout, and Andras Szigeti, so many thanks to them. I am also independently grateful to Chandra Sripada for urging me to write this essay in the first place. I made many revisions to earlier drafts of this essay while on academic leave, on a subgrant from the John Templeton Foundation for Alfred Mele's project on the Philosophy and Science of Self-Control. I am very grateful for this financial support, as well as for the leave provided by my home institution, Tulane University.

1. As Gary Watson (2014, 15–16) rightly notes, while few articles have received as much attention as "Freedom and Resentment," "very few philosophers embrace its fundamental program," which includes a "'response-dependence' thesis" without which "we have nothing close to a Strawsonian understanding of responsibility."

Philosophical Review, Vol. 126, No. 4, 2017
DOI 10.1215/00318108-4173422

function of being held responsible, that is, it is somehow a function of being a target of such responses.

This feature of Strawson's approach has struck many responsibility theorists as deeply implausible. It is nevertheless the position I intend to argue for and defend here. It will be, as far as I can tell, the first attempt to provide an actual argument for this view.[2] Strawson himself just states the position, rather obscurely at that, and his characterization lends itself well to misinterpretations. On his picture, fundamental to our nature is a kind of sociality, and given our sociality, it matters greatly to us what intentions and attitudes others have toward us. Consequently, "we demand some degree of goodwill or regard on the part of those who stand in these relationships to us" (Strawson 2003, 76). The reactive attitudes are the natural responses that we have to those who violate this demand (albeit with excuses for some whose injurious behavior doesn't exhibit ill will, and exemptions for those who are incapacitated in some respect "for ordinary interpersonal relations" [Strawson 2003, 82]). But as the demand and our emotional responses to its violation are grounded in what we care about, these responses are what determine the nature and contours of moral responsibility (and so help fill in and ground our practices of moral condemnation and punishment), and as this general system of responsibility responses structures our form of interpersonal life, there simply is no question as to its correctness or incorrectness from an external standpoint (Strawson 2003, 90–92).

As I said, this view is more or less just asserted in "Freedom and Resentment." But most theorists have thought the contrary view is what is obviously true instead, that holding responsible is instead a function of *being* responsible: resentment for an injury, for example, could be appropriate only in light of the injuring agent's antecedent responsibility for the injury. We see this thought expressed by Fischer and Ravizza (1993), for example, who question Strawson's response-dependent view (at least in its most naive form) by noting that a world in which minorities and

2. Others who have provided a sympathetic interpretation of Strawson, including Bennett (1980) and Watson (2004, 219–59; 2014), essentially fill in or clarify some of the more obscure aspects of Strawson's picture, but they don't really fill in any argumentative details, at least regarding the response-dependent aspect of the account. Allan Gibbard (1990, 40–45) comes closest when offering an argument about the relation between anger and the blameworthy, but his primary aim is to develop an *analysis* of the terms in question, and in particular the term "rationality" as it pertains to "ought" and "morally wrong." Nevertheless, I believe he would be quite sympathetic to the treatment I am about to give.

women were not held responsible or a world in which those with profound intellectual disabilities were held responsible would be worlds whose members' responsibility attitudes and practices were just mistaken. As they put it, "By understanding responsibility primarily in terms of our actual practices of adopting or not adopting certain attitudes toward agents, Strawson's theory risks blurring the difference between" being responsible and holding responsible (Fischer and Ravizza 1993, 18).[3]

The much more popular alternative view, then, is a response-*independent* view of responsibility, according to which there are antecedent properties of being responsible that our practices of holding responsible must respect and respond *to*. On this view, our reactive attitudes make sense only if we presuppose the responsibility of the target. Consequently, "The reactive attitudes are *evidence* about when to hold people responsible, but not something that constitutes them being responsible" (Brink and Nelkin 2013, 287; emphasis in original; also a view they call "realist"; see also Arpaly 2006, 28, 31).

I admit that the response-independent view of the nature of responsibility seems most natural. Nevertheless, it is false. I obviously have the burden of proof here, though, so in what follows, I will first attempt to mount a burden-shifting argument in favor of my view (drawing from some recent insightful work in metaethics), and then I will spend the remainder of my time defending the view from a host of objections. My positive argument depends on an extended analogy with humor: One of our paradigm responsibility responses is exactly like our paradigm humor response in many surprising respects, and as the most plausible theory of the nature of humor is a response-dependent one, so too is the most plausible theory of at least one familiar domain of responsibility.[4] The key to making the case for response-dependence in

3. See also Todd 2016.

4. Patrick Todd (2016, 236–38) has also recently explored the analogy between response-dependence about the funny and response-dependence about responsibility. However, he doesn't consider my version of the analogy, which rests on a thoroughly normative response-dependent account of both. As just implied in the text, I will try to defend a response-dependent view of the funny that doesn't succumb to the worry that if, say, we were all to find genocide funny that it would thereby be funny. What I am exploring is thus what Todd calls a mysterious "middle ground" approach that he says would be a "neat trick, if it could be done," something nevertheless "eminently worth trying to do" (Todd 2016, 237). This general approach has been available for a while in the sentimentalist tradition, though, explicated most recently and fully in a series of papers by Justin D'Arms and Daniel Jacobson (especially D'Arms and Jacobson 2006) that I will discuss and cite below.

both humor and responsibility is to recognize that there is no plausible way to identify the values regulating the emotional responses we have to an incredibly diverse range of events and activities in each domain unless we make essential constitutive reference to those emotional responses. After fending off numerous skeptical objections to my approach, I will conclude by exploring some implications of the view for our theorizing about other domains of responsibility.

Part 1: The Funny

The Funny List

What do the following items have in common?

- An *Onion* headline in 1998: "God Answers Prayers of Paralyzed Little Boy. 'No', says God."
- A dead-on impression by your friend of Christopher Walken saying, "I gotta have more cowbell!"
- "Your momma is so fat, that when she lies around the house, she lies *around* the house!"
- Steve Martin suddenly and deliriously dancing around while exclaiming, "Oh no, I have *happy feet*!"
- Laurel and Hardy trying and repeatedly failing to move a piano up a long flight of stairs
- One cartoon character dropping an anvil onto the head of another
- A squeezed ketchup bottle producing a farting noise
- Two six-year-old boys laughing hysterically at a squeezed ketchup bottle producing a farting noise
- Monty Python's "Ministry of Silly Walks"
- Jonathan Swift's "A Modest Proposal"

When we look at these items with an objective eye, they constitute quite a strange and diverse group, one that includes violence, sacrilege, pain, bodily emissions, delirious joy, balletic clumsiness, silliness, ordinary age-specific behavior, ambiguous language, mimicry, irony, insults, and incompetence. But when we look at them from our engaged human perspective, there are two obvious ways to see the group as perfectly unified. First, these are all things to which we tend to respond with amusement. Second, they are all things that are funny. Here, then, is our question: What is the precise relation between our amusement and the funny? There are three possible answers.

Dispositional Response-Dependence

The first answer is analogous to a caricature version of Strawsonian response-dependence, appealing to how people in fact respond to all of these items:

> **Dispositional Response-Dependence about the Funny**: The funny is what typically elicits amusement; that is, something is funny to the extent that people are typically disposed to respond to it with amusement, under standard viewing conditions.

So what unites all the items on the list as funny is simply that people are in fact disposed, under a certain set of privileged conditions, to be amused by them.[5]

This view is quite implausible, however, for three reasons. First, the reasons to which we refer in affirming or denying that something is funny never (or extremely rarely) make reference to people's dispositions; rather, they make reference primarily to properties of the purportedly amusing object. Imagine how jarring it would be to hear someone say that the real reason the comedian Carrot Top is so funny is that so many people find him amusing. Indeed, second (and more importantly), we seem quite committed in our evaluative interactions with one another about the funny to the regular possibility that people's actual, typical amused responses (even in standard conditions) may be wrong. "I don't care how many people find Carrot Top funny," we might say, "he's just not funny!" Third, a plausible dispositional treatment of the funny would need to provide us with a substantial account of the relevant "standard" conditions that privilege a certain set of amused responses as those that count in determining the contours of the funny. But it is quite difficult to imagine what those might be (see, by way of comparison, D'Arms and Jacobson 2006, 201).

Consequently, we should reject the dispositional version of a response-dependent analysis of the funny in favor of an account that incorporates normative standards. In doing so, many theorists have embraced a response-independent account. Let us examine that possibility.

5. This seems to be the only non-Platonic view of the funny that Todd (2016) thinks there is.

Response-Independence

We are, in our interpersonal lives, clearly committed to normative standards of the funny beyond just those established by what typically amuses people. As Jacobson (2011) notes, "To call something funny is in some way to endorse amusement at it, not to report or predict it" (emphasis in original). Something is funny if and only if it merits amusement. But there are two ways to read this biconditional, depending on which has priority, the funny or our amusement. In this section, I explore the response-independent reading.

> **Response-Independence about the Funny**: The funny consists in a property of objects that makes amusement at the objects appropriate, a property whose funny-making status is ultimately independent of, and grounds, any actual amused responses to it. Something thus merits amusement if and only if, *and in virtue of the fact that*, it is antecedently funny.

Before explication, a quick clarification is in order. Often when we are theorizing about what is funny, we will test a theory by appealing to whether or not we are amused by some purported counterexample. To appeal to our amusement in this way is to grant our responses *epistemic priority* in our theorizing (McKenna 2012, 39–42). That sounds like response-dependence of a sort, but for our purposes it is not the right sort. Why, after all, can we appeal to our responses to do this epistemic work for us? It is because, many think, our responses tend to reliably track the truth about the funny. On this construal, then, the funny is *metaphysically* prior to our amused responses. What our responses track (when properly functioning) is what is antecedently funny, and so what merits amusement is just a function of the funny. It is that metaphysical sense of response-independence about the funny that we are concerned with here.

If the value in question—the funny—is ultimately response-independent, then the advocate of this view owes us an account of what properties other than our amused responses that value consists in. In other words, what is it that unifies the items on the funny list as funny, and so properly regulates our amusement at them? Answering this question has been the aim of many theorists over the years. I will discuss only the two leading contenders. According to the *Incongruity Theory* (Clark 1987; Oring 2003), the funny consists in the incongruity between our expectations and our experiences. According to the *Benign Violation Theory* (McGraw and Warren 2010; McGraw, Williams, and Warren 2013), the funny consists in the benign violation of a norm.

Both theories purport to provide a criterion of the funny, putting forward something as funny if and only if, *and because*, it involves incongruity or benign norm violation. But there are prima facie counterexamples to each theory, some drawn from various items on our funny list. Consider false negatives. There seems nothing incongruous, nor does there seem to be a benign norm violation, when a six-year-old boy laughs uproariously at his own fart (the fart may be incongruous or a norm violation, but the laughter isn't; that is just what six-year-old boys *do*). Where is the incongruity or norm violation in a piano that slips from someone's grip and goes crashing down the stairs? Now consider false positives. The incongruous sometimes produces only bafflement (a Dali painting), and benign norm violations are sometimes merely naughty (getting caught in a white lie) or nothing of note (when one is too tired to floss one's teeth).

Now one might attempt to tweak the theory to avoid these prima facie counterexamples, but how? As I will suggest below (when talking about shadow skepticism, a more sophisticated version of response-independence), the only way to do so is to restrict the relevant incongruity or benign norm violation, say, to the things that are incongruous or violations *in an amusing way*.

A response-independent theory of the funny must capture all and only the funny, identifying the property that regulates and unifies our amused responses without constitutive reference to those responses. Neither of the leading theories does so without easy counterexamples, and it is hard to imagine any other response-independent theory doing better; hard to imagine, that is, how any attitude-independent feature(s) of the world could impose a unified grouping on "the funny" (Scruton 1987; Wiggins 1987, 193; D'Arms and Jacobson 2006, 194–96; Jacobson 2011).[6] So is there a better alternative?

The Burden-Shifting Argument

The better explanation of what unites all the items on the funny list is just that *they are all things to which humans appropriately respond with amusement.* This is a response-dependent theory of humor with normative bite. But it

6. Perhaps a paradigm-case, family-resemblance approach would work instead? The problem with this approach stems directly from the funny list again: What kind of weird, dysfunctional and disjointed family *is* this? What makes *A Modest Proposal* anything like Steve Martin's happy feet, other than that they are all things that we find amusing?

needs quite a bit of explanation and even more defense. Let's start with an official statement of the view:

> **Fitting Response-Dependence about the Funny**: The funny *just is* whatever merits amusement (the amusing). That is to say: something is funny if and only if, *and in virtue of the fact that*, it merits amusement.

On this view, amusement is more metaphysically basic than the funny. But note that it is not amusement per se that is more basic; rather, it is *merited* amusement. Merit is most often cashed out in the literature in terms of *fittingness* (see, for example, D'Arms and Jacobson 2000, and forthcoming; Rabinowicz and Rønnow-Rasmussen 2004; Bykvist 2009). As a pancultural emotion, amusement has a syndrome with three key features: It is (a) an irruptive affective state that (b) appraises some object as amusing, and (c) is generally thought to be identified and differentiated in terms of its *action tendency*, which is obviously to laugh (Frijda 1986, 2007; Scarantino 2014). Of course, this tendency can be squelched, as one is supposed to do at funerals and in church. But its readying us to laugh is the main thing that differentiates amusement from other positive emotional states like, say, admiration, that may have a similar phenomenal feel. Now one has a *reason of fit* to feel amused at something just in case that object has the properties that amusement appraises it as having (D'Arms and Jacobson 2000). So what unifies the items on the funny list, on the fitting response-dependent view, is just that it would be fitting to respond to each of them with a pleasant affective state whose action tendency is laughter and that appraises them as amusing.

I want to be as clear as possible about what is going on here, as doing so will shed light on a more subtle issue when it comes to responsibility. There are three items we have mentioned: amusement, the amusement-worthy (that is, what merits amusement, or what amusement fits), and the funny. The funny is the value that we are trying to get a bead on. The dispositional response-dependent view discusses only two of these items, the funny and our amusement, and it maintains that the funny is just whatever in fact amuses us (under standard conditions). What is missing from this account is normativity, and that is what our third item—the amusement-worthy—imports. Our quest has then become to find the relation between the funny and the amusement-worthy. The response-independent account maintains that the funny consists in some response-independent property (for example, incongruity or benign norm violation) that subsequently makes amusement worthy. The fitting response-dependent view, by contrast, says that what makes

something funny *just is* that it is amusement-worthy. The value that "funny" picks out is wholly captured by whatever it is that amusement fits, and in virtue of amusement's fit (see, by way of comparison, D'Arms and Jacobson 2006).

So what are the fittingness conditions of amusement, that is, what makes amusement fitting? This is an ambiguous question. On the one hand, it might refer to what, *in any individual purported instance of the funny*, makes amusement fitting. Here the answer will refer to specific objective properties of the appraised target. We may then disagree about whether such properties are actually instantiated in this object, but this is a boring sort of empirical disagreement, and so a boring version of the question. This version of the question is also neutral with respect to the response-dependent/-independent debate about the funny, for both sides can agree that there are objective (mind- and attitude-independent) properties of funny objects to which amusement appropriately responds.

Thus the type of disagreement dividing response-independent and fitting response-dependent theorists has to be over what it is that makes certain objective properties *count* as being in the class of things to which it is fitting to respond with amusement in the first place. To return to one leading theory, there are of course instances in which incongruity is funny, and so it will be fitting in those instances to respond with amusement to it (D'Arms and Jacobson 2006, 194). The interesting sense of our question, though, asks why incongruity in these instances counts as a fitting-amusement-maker? And here is where the true fitting response-dependent answer shows up: it is simply the type of thing to which our humor sensibilities, in those instances, are built to respond with amusement.

Deep and interesting disagreement about the funny is about what properties our human sensibilities have rendered the fitting-amusement-makers, and this disagreement may be irresolvable (D'Arms and Jacobson 2010, 606). People do have different comic sensibilities. At this level of disagreement, all we can do is urge the other person to look harder, to be more open, to try and see it in the way we do, to avoid inconsistency, and so forth. But before abandoning the possibility of resolution, we can try to make sure nothing is obscuring our vision of the comic properties. For example, perhaps one of the parties to the disagreement is depressed, exhausted, or has heard the joke one too many times to be amused. And it is also possible that some people just have defective senses of humor. Perhaps those raised on overly sensitive children's TV programming, or those who are too nice or too easily offended, cannot see their way into

being amused by dark humor, for example.[7] People also have deficient sensibilities in other arenas (music, food, art). It is a flaw in their human machinery, albeit often a correctible flaw. So while we may not often be able to resolve the most fundamental disagreement about the funny-makers, that doesn't mean that in many cases of disagreement one side isn't right.

That is the fitting response-dependent account of the funny in a nutshell. The prima facie burden-shifting argument in its favor is that the most plausible unifying account of the items on the funny list is that they are all fitting targets of amusement (amusement-worthy), where the fittingness of this response is ultimately determined by our properly firing comic sensibilities. The value in question, the funny, *just is* the amusement-worthy. Now this is a rather modest argument on its own, but the plausibility of its conclusion will increase quite a bit if we can mount a defense against its most powerful rival.

Shadow Skepticism

There are several skeptical worries one might have about a fitting response-dependent theory of responsibility, as we will see, but there is only one real skeptical worry about the analogous theory of the funny worth taking seriously. This is what D'Arms and Jacobson (forthcoming) call *shadow skepticism.*[8] It is a more sophisticated version of the original response-independent theory. The basic thought is that response-dependent "values" actually shadow response-independent properties. To get a feel for the position, consider the natural emotions of fear and disgust. A response-dependent treatment of them would say that the associated values are the *fearsome* (whatever merits fear) and the *disgusting* (whatever merits disgust). But the shadow skeptic says that both emotions are clearly the product of natural selection, having evolved to respond to objective properties of the environment, respectively, *the dangerous* and *the contaminated.* While fear and disgust may therefore be rough-and-ready guides to those properties, the properties themselves are what constitute the real values in this vicinity. This means that the fearsome and the disgusting must be either identical to those properties or they really don't capture what matters to us (which are those properties). In other words,

7. Thanks to Daniel Jacobson for discussion.

8. The objection is one against sentimentalism about value generally, which I'm here applying specifically to fitting response-dependence about the funny.

the fearsome and the disgusting are mere shadows of, respectively, the dangerous and the contaminated, and it is those latter response-independent properties that we actually value.[9]

So too, the thought might go, amusement evolved to track various response-independent features of the environment, and those are the properties that matter to us, which means "the amusement-worthy" is actually either identical to those response-independent properties or, if not, it is not what really matters us. Our amusement, then, is at most an epistemic tracker of the funny, not a metaphysical constituent of it.

Of course, the evolutionary stories of our fear and disgust are much more obvious and easy to tell than an evolutionary story of our amusement. Nevertheless, let us consider one theory that has gained traction just to see how we might respond. It is the *play theory* (Van Hooff 1972; Pinker 1997). Playing—engaging in it and anticipating it—produces pleasure (that is, dopamine), and play-deprived animals evince all sorts of social dysfunctions. What is the value of play? It is thought to train people both for the expected (for example, what to expect in throwing some object—a stone, javelin, or football—at some target) and the unexpected (in skipping, skateboarding, or jumping in unusual ways we test the limits of our balance and control) (Spinka, Newberry, and Bekoff 2001). We are training, in other words, for evading predators and hunting. Lots of this training in early years involves roughhousing. But it is important that play fighting does not become *real* fighting. Consequently, we have developed a social "play face," which consists in smiling or laughing. It indicates to others that we are "just playing" (Van Hooff and Preuschoft 2003; Boyd 2004, 8–9). And the connection of laughter to expectations yields the evolutionary hypothesis that laughter (and amusement generally) has come to be a response to *surprise*, that is, being caught off guard by something, but in a way that quickly allows us to recover our senses, to learn or appreciate something new going forward, and to reinforce our sense of solidarity with one another (Boyd 2004, 16).

The basic thought of the shadow skeptic, then, would be that the value that organizes and unifies all the items on our funny list is actually *solidarity-reinforcing playful surprise*. This would be the response-indepen-

9. See D'Arms and Jacobson's helpful articulation and discussion of a version of this argument by Schroeter (2006) on PEA Soup (blog), February 14, 2014, http://peasoup.typepad.com/peasoup/2014/02/featured-philosophers-darms-and-jacobson.html.

dent property that we purportedly care about, the property that the amusement-worthy merely shadows.

There are two powerful replies to this thought. The first, adapted from Scruton (1987) and D'Arms and Jacobson (2006 and forthcoming, chap. 7), is that, even if correct, the identified value is actually covertly response-dependent. To see why, ask yourself how we should understand "playful surprise." The hard question is what "playful" consists in, as that is the key amendment for fending off easy counterexamples to a purely surprise-based theory; after all, the surprises must be *treats*, not threats (Boyd 2004, 12). Now, there may at first seem to be plenty of false negatives, given that "playful" seems to connote intention yet lots of funny things are accidents, for example, Laurel and Hardy's bumbling piano movers, a fart, or a malapropism. But suppose we remove intention (and so incorporate these cases) by thinking of "playful" merely as *nonthreatening*. Benign surprises, if you will. Now it looks like our false negatives will be significantly reduced, if not eliminated.

But what about false *positives*, that is, benign surprises that aren't funny? There seem to be many: a twist ending in a drama, hearing the doorbell while being lost in thought, a new guitar riff from your bandmate, drinking orange juice when expecting milk. Perhaps, then, we mean *delightful* surprises? But again, there are plenty of false positives: seeing a loved one who has unexpectedly returned home from a tour of duty; being engrossed by a dinner guest who turns out to be an engaging raconteur; winning the lottery.

In terms of being a playful surprise, there seems to be no difference between my shaving my head while my partner is away for the day and my shaving merely *one-half* of my longtime bushy head (the front half, say). But only the latter is funny. Why? Only the latter involves shaving my head *in an amusing way*. But that is just to make essential reference to response-dependence. To the extent that the funny does involve playful surprise (what I merely granted for the sake of argument here), it would be only the subset of playful surprises *meriting amusement* that could adequately characterize the value in question.

A second reply to the shadow skeptic goes to the relevant order of causation in the evolutionary story. The shadow skeptic's line suggests that we evolved to be amused at playful surprises (or whatever the relevant response-independent property is) *because* doing so promoted our ancestors' reproductive success. That is, there must have been some valuable property of various events—playful surprise—that some early humans responded to with amusement, which then increased their fitness in such

a way that we have come to inherit that trait. But this seems the wrong causal story. Rather, it is much more likely that responding with amusement and laughter to some property made others (and the laugher) behave in a way that increased the laugher's (and her descendants') own fitness; that is, perhaps the laughter caused pleasure or some tendency to cooperate (or play), and *that's* why the trait generated fitness and was inherited by us. But then there would have been no tracking of response-independent funny properties occurring at all.

To conclude part 1, then, I think there is a clear and powerful burden-shifting argument in favor of the fitting response-dependent view of the funny, an argument whose plausibility is then bolstered by two good responses to the shadow skeptic. But so what? I have deliberately spent a lot of time developing and defending this view for a simple reason: I believe that the domain of humor is tightly analogous to a significant domain of responsibility, and so we can appeal to moves made with respect to the less controversial former theory to make sense of, articulate, and defend the much more controversial latter theory.

Part 2: Responsibility

The Wide Range of Responsibility Responses

Before exploring the analogy between humor and responsibility, I must address a complication: while there is basically just one emotional response appraising the funny—amusement—there are many distinct emotional responses appraising "the" responsible. Strawson (2003) primarily discussed three—resentment, indignation, and guilt—but he mentioned several more, including gratitude, hurt feelings, forgiveness, and love. There are many more responsibility responses than that, though, including admiration, disdain, disesteem, contempt, abhorrence, hatred, revulsion, shame, regret, pride, disapproval, disappointment, irritation, frustration, approval, appreciation, approbation, gratification, and warm feelings (Shoemaker 2015, 35). Note that there are positive and negative responses on this list, and some have very different targets (for example, character, judgments, actions, and attitudes), a fact some have thought may implicate different types of responsibility (Watson 2004, 260–88; McKenna 2012, 7–9; Shoemaker 2015; Pereboom 2014, 2017). Nevertheless, I will focus exclusively here on what nearly everyone agrees is a type of responsibility, what is usually called *accountability*. The reason most theorists take it to be a significant part of the responsibility terrain is that its negative form—being accountable for something bad—is just

what tends to conjure up Strawson's titular reactive attitude, *resentment* (see, for example, Wallace 1994; Fischer and Ravizza 1998; Watson 2004, 278; Wolf 2011; McKenna 2012, 2–3), and, further, this attitude tends to involve sanctions, which cry out for justification (see, for example, Watson 2004, 278; McKenna 2012, chap. 7; Pereboom 2014, 178–86).[10]

While I will focus on negative accountability in what follows, I will not, perhaps surprisingly, be discussing resentment, primarily because it no longer means in the philosophical literature what Strawson actually meant by it, and leaning on the meaning it has come to have would beg the very question at issue. What Strawson had in mind was merely a natural angry response to a discernment of another's ill will (or insufficiently good will; Strawson 2003, 80; Deigh 2011). His notion of the emotion thus seemed to be noncognitive. Nevertheless, "resentment" has come for many theorists to include a constitutive cognitive component, typically a judgment that the offender either wronged one or did something blameworthy to one (see Shoemaker 2015, 88n1 for a representative list of such theorists). But if that is the right characterization, and resentment is the paradigmatically appropriate emotional responsibility response in the domain of negative accountability, then it seems clear that it (partially) consists in a judgment of the target's *antecedent responsibility*. But that characterization of resentment makes responsibility response-independent from the get-go. Instead, then, given that everyone at least agrees that resentment's (and indignation's) purely emotional core is *anger*, I focus here just on that emotion so as not to beg any questions (see Shoemaker 2015, 87–91 for more discussion on this crucial point; see also Gibbard 1990, 129–32; and D'Arms and Jacobson 2003 for independent arguments against there being a judgmental component in anger). What I will discuss is the very familiar kind of anger that we tend to experience toward other agents, an emotion which involves feelings of heat and aggression, appraises an agent negatively in light of some perceived bad action or attitude, and includes a motivational impulse (an action tendency) to communicate itself to the offending party (often via retaliation) (Shoemaker 2015, chap. 3, and forthcoming). At the end of the essay, I will gesticulate a bit about whether and how we might apply the response-dependent treatment to several

10. I strongly disagree with the thought, though, that there is any kind of necessary connection between the angry blaming attitude and sanctions. I discuss this idea in Shoemaker 2015, 103–12; forthcoming.

other emotional responsibility responses as well, a task whose path will have been thoroughly laid out by then.

The Blameworthy List

What do the following items have in common?

- A Yankees fan at a Bronx bar punches you in the face because you are wearing a Red Sox cap.
- You read a story about a priest who publicly confesses to having sexually molested several altar boys.
- You see someone at a political rally, who, to show his contempt for some governmental policy, repeatedly farts and spits on your country's flag, a country (and flag) you fought and were wounded for during a war.
- You spy your best friend mocking your physical disability at a party.
- You read a story in the paper about a man who drove to work and left his car in the hot sun all day, having forgotten that his infant was in the back seat.
- You stumble across a diary entry of your spouse talking about having had a secret affair.
- In presenting your philosophical research at a conference, you notice that a big shot in the audience is rolling his eyes and shaking his head, mouthing the word "idiotic."
- You witness a parent in a store slapping her child.
- You hear a man in a movie theater laughing every time a woman in the film is beaten by her husband.
- An employee of yours does everything you ask, but with obvious condescension.
- A workman tosses pieces of heavy slate off a roof to the ground below without checking to see if anyone is there, and he just misses hitting you as you walk by (Hart 2008, 147).
- Your spouse promises to pick up some milk on her way home from work, but because she becomes lost in thought about a paper she is writing, she forgets to do so (see Clarke 2014).
- A friend comes to visit you in the hospital, but when thanked, she responds, "I only came because it is my Christian duty."
- You find out that your jealous roommate deliberately omitted to tell you that the person you have been hoping would call and ask you out actually did so last week.
- As one texting driver passes in front of your house, you manage to contain your squirmy dog, but as another texting driver gets near, your dog jumps out of your arms, runs into the street, and that second texting driver hits him.

This is, as with the funny list in part 1, a very odd group. It includes, as did the funny list, examples of violence, pain, bodily emissions, clumsiness, obliviousness, stupidity, and incompetence. But instead of arousing amusement, these are all items that tend to arouse anger. We also tend to think that these are items for which the targeted agents are negatively accountable, that is, they are *blameworthy*.[11] Our question, then, is this: What is the precise relationship between anger and the blameworthy?[12] As with our amusement and the funny, there are three possible answers.

Dispositional Response-Dependence

The first answer involves a flat-footed reading of Strawson, a position we can call

> **Dispositional Response-Dependence about the Blameworthy**: The blameworthy is what typically elicits anger; that is, it is what people are typically disposed to respond to with anger, under standard conditions.[13]

So what makes all the agents on the above list blameworthy, on this construal, is that people are disposed, under a certain set of privileged conditions, to be angry at them.

11. I fully appreciate that there are other sorts of blameworthiness, including *aretaic*. But I will be using "blameworthy" in this essay to refer to worthiness for the kind of response that we take to be aptly associated with negative accountability (whatever that is).

12. Why think that we can gain an understanding of the blameworthiness of negative accountability via a survey of anger? After all, this isn't something that we tend to feel all the time in responding to accountable people who do something bad, and if we did, we would need serious therapy. In addition, we have all sorts of nonangry negative accountability responses, so why privilege anger in gaining access to accountability? (Thanks to an anonymous referee for raising these concerns.) One reason is historical: I am trying to vindicate Strawson's response-dependent treatment of responsibility, which is for him, on the negative side, the domain of resentment, indignation, and guilt, all forms of anger (and which differ for him only in terms of the relation between the blamer and the offender). Another reason has to do with theoretical ecumenism: I am trying to capture what most accountability theorists do take to be the apt response to negative accountability (see the citations in the text). Finally, even if we do have other negative emotional responses in our accountability arsenal, anger does seem to be our go-to paradigm response. However, it is crucial to point out that admitting these points doesn't mean that we *display* or *express* this sort of anger all the time (or even very much). Rather, my focus here is only on the appropriateness of *feeling* anger as a negative accountability response. Its display and proper expressions, I maintain, are matters of ethics, not responsibility (see Shoemaker 2015, 221–23).

13. Again, those theorists who seem to have interpreted Strawson dispositionally like this include Bennett (1980), Fischer and Ravizza (1993, 16–19), and Todd (2016).

The same three reasons against dispositional response-dependence about the funny apply here even more forcefully. First, the reasons to which we refer in judging someone blameworthy just do not make justifying reference to anyone's dispositions; rather, they make reference to features of the blamed agent: "You stepped on my foot!" or "You broke your promise!!" or "You're drunk again?!" Imagine the following bizarre complaint instead: "I deem you blameworthy because most people would in fact be angry at you."[14]

Second, and more importantly, people's typical angry responses, even in standard conditions, may be *wrong*. This is precisely the point made against a purely dispositional response-dependent reading of Strawson by Fischer and Ravizza when they consider what we would say about a community whose members all resented the profoundly intellectually disabled or refused to resent women and minorities. As they put it, "Surely it seems possible that one can be held responsible even though one in fact is not responsible, and conversely that one can be responsible even though one is actually not treated as a responsible agent" (Fischer and Ravizza 1993, 18).

Finally, what could be the "standard conditions" for responding with anger that would privilege some people's actual angry responses over others in determining the blameworthy? It is hard to know even where to begin to answer this question.

Theorists have come to agree, therefore, that we need a genuinely normative characterization of the angry responses associated with the blameworthy. As Fischer and Ravizza (1993, 18) put it: "Agents are morally responsible if and only if they are *appropriate* recipients [of] the reactive attitudes." Someone is blameworthy, on this construal, if and only if anger at her is appropriate, what I will label *angerworthy*. There are two ways to read this biconditional, however, depending on which has metaphysical priority, the angerworthy or the blameworthy.

14. It is true, of course, that sometimes we do say things like, "Most other people would be angry too if they were in my position!" (Thanks to an anonymous referee for noting this point.) This is typically offered, it seems to me, as cover for having blaming anger, not as an appeal *justifying* having it, that is, an appeal to what makes the target blameworthy. *Fifty Million Elvis Fans Can't Be Wrong*! was the informal name of Elvis Presley's ninth record, and this boast congratulated those who recognized Elvis's great-making properties as being in popular company, but his singing properties were what allegedly made the record worth buying, not the fact of the mass purchase.

Response-Independence

The first way to read the biconditional is response-independent:

> **Response-Independence about the Blameworthy**: The blameworthy consists in a property (or properties) of agents that makes anger at them appropriate, a property (or properties) whose value-making is ultimately independent of our angry responses. Anger at someone for X is appropriate if and only if, *and in virtue of the fact that*, she is antecedently blameworthy (and so accountable) for X. What makes her blameworthy is thus ultimately response-independent.

If you favor this way of putting the relation between being and holding responsible (at least with respect to anger and the blameworthy), believing that being responsible is metaphysically prior to holding responsible, then you owe us an account of what the blameworthy property is, an account in which anger is, at most, a reliable epistemic tracker of the blameworthy. But just as with the response-independent theories of the funny, response-independent theorists of blameworthiness are, I will argue, vulnerable to serious prima facie counterexamples no matter which way they turn.

We will be attempting to answer the following question: *What makes an agent blameworthy for something bad (B) independently of any reference to the appropriateness of anger?* What follows are some familiar possibilities. The most popular contemporary view will be constructed from attempts to counter some obvious false positives to the first three conditions.

1. ***B* was generated *voluntarily*.** Perhaps all that matters for blameworthiness is that one chose to do the bad thing, that it was the product of one's actual will. This was, more or less, how Hobbes and Locke viewed the matter. However, having this property is insufficient for blameworthiness, as illustrated by those who act on their compulsive desires, manias, or addictions (see, for example, Frankfurt 1988, 11–25; Watson 2004, 13–32). These agents do what they in fact choose to do, in some sense, but they lack a key modal property, namely, the ability to refrain from those choices or activities, or to be responsive to alternative reasons. We thus need to add such a property to account for these agents.
2. ***B* was generated voluntarily and under *control*.** The long-presumed connection between free will and moral responsibility has been taken to be a matter of determining the maximum amount of control necessary for moral responsibility (see, for example, Fischer and Ravizza 1998, 28–34; Mele 2006; McKenna 2006, 24; McKenna

2012, 10). There have been numerous explications of the type of control thought necessary: mechanism-based reasons-responsiveness, agential reasons-responsiveness, a libertarian ability to do or choose otherwise, having ultimate sourcehood, and more. But no matter how one fills in this story, voluntary control alone is still insufficient for blameworthiness. In illustration, suppose I voluntarily, and under control, fix you a cocktail, but I do not know that someone has replaced my rye with poison. I am not blameworthy for poisoning you. We thus need to add another familiar condition to our account.

3. ***B* was generated voluntarily, under control, and *knowingly*.** We may fill in the knowledge component, like the control and voluntariness components, in various ways. It is standardly thought to involve both knowledge of what one is doing and knowledge that what one is doing is wrong/bad (see Vargas 2013, chap. 7). So if I know that there is poison in the rye bottle, and I know that poisoning you is wrong, and yet I voluntarily and under control fix you a cocktail from it, I am blameworthy for poisoning you.

And that is where most response-independent theorists leave it: *An agent is blameworthy—and so as a result merits anger—for something bad if and only if, and because, the agent generated it knowingly, voluntarily, and under control.*[15] Nevertheless, we can find prima facie counterexamples to every single aspect of this formulation. I will begin by discussing false negatives for each condition individually, and then I will discuss false positives for the set.

Knowledge

Start with the epistemic condition. When the worker tosses the slate off the roof without thinking and almost hits you as you walk below, he is surely blameworthy. But he is ignorant (at the time) of the fact that what he is doing is wrong (see Clarke 2014, 160). This is classic negligence. He should have taken a simple precaution, looking to see if anyone was below. But he didn't.

15. I take these three ways, more or less, to match up with Brink and Nelkin's thorough treatment of the architecture of responsibility, where they argue that what is required for (realist, response-independent) responsible agency is normative competence (which requires both epistemic and volitional capacities) and situational control (which, when lacking, is what excuses some cases of compulsion, constraint, coercion, and duress). See Brink and Nelkin 2013, 292–303.

Now suppose that his ignorance here resulted from some *prior* blameworthy act; for example, he took a pill an hour ago to deliberately cause himself to be oblivious in working on the roof (perhaps so that he could be more efficient). This would make his negligence culpable, which a lot of people think is required for him to be blameworthy for his failure now (Rosen [2003, 64; 2004, 300; 2008, 600], Zimmerman [1997, 414; 2008, 175], H. Smith [1983, 2011], FitzPatrick [2008]). They then add this as an extra clause within the epistemic requirement above: One must know that what one is doing is wrong at the time of the action, *unless one's ignorance of that fact is culpable*. But allow what is much more likely to be the case, namely, that our worker was *not* culpable in this way: he was at no time aware that anything he was doing would lead to this result. He was just oblivious all along. Still, he seems blameworthy, the appropriate target of your anger when a piece nearly hits you (Clarke 2014, 168–69).

Now there are some who deny this result. They say that if there is no way to trace back what he did to some prior (knowing, voluntary, controlled) bad action that is the source of his current ignorance, then he just cannot be blameworthy for it (see Zimmerman 1997, 414; Rosen 2004; and Levy 2011, chap. 5). This is the view I will call *theory-drenched skepticism*, and I will deal with it in detail later. Let us focus now instead on nonskeptical alternatives, those that accord more pride of place to our emotional responses. After all, one might say, anger seems just as appropriate in some cases of nonculpable ignorance as it does in cases where knowledge is in place, so perhaps we simply need to tweak the knowledge condition some more to account for them. What is necessary instead, many suggest, is either that the agent knew what he was doing was wrong *or that he should have known* it was wrong, where the latter simply amounts to there being an applicable cognitive standard (relative to the agent and his situation) below which the agent fell (see, for example, Sher 2009; Clarke 2014, 167).

Notice, first, that this is no mere "tweak" of the knowledge condition; rather, it is an admission that knowledge is actually *unnecessary* for blameworthiness, and so what it introduces is a hybrid epistemic condition for blameworthiness, filling in what is required when actual knowledge is absent. This move also assumes that "ought implies can," that if one should have known, one *could have* known (Clarke 2014, 167). So in a case of blameworthy ignorance (as in the slate tosser), there is nevertheless an epistemic standard the agent could have met. But what is that standard precisely, and, more importantly, what determines it?

In order to be plausible, the standard has to be relative to the specific capacities of individuals as well as to their specific circumstances (Clarke 2014, 102–4). A nongenius or an outsider cannot be blameworthy for not knowing things only a genius or an insider could have known. And on the flip side, someone deprived of moral knowledge throughout his upbringing cannot be expected to know what the rest of us do about morality. If someone has been raised in a thoroughly racist household in the isolated bayous of Louisiana, he cannot be expected to know that racism is wrong, and so he cannot be blameworthy for favoring only his "own kind" or treating those of other races poorly.

Except that he can be. Or at least that is the judgment of an overwhelming number of respondents when such cases are empirically investigated, shown repeatedly by my coauthor David Faraci and me (Faraci and Shoemaker 2010, 2014, and n.d.). We were interested in the extent to which moral ignorance due to childhood deprivation affects attributions of blameworthiness. What we found (with very little variation over three different studies on a large and diverse set of subjects) is that moral ignorance of that sort mitigates people's attributions of blameworthiness to the wrongdoer a bit, but that people nevertheless find him to be seriously blameworthy.[16] On the basis of what? Perhaps, subjects think, even though our agent didn't know beating peasants on a whim was wrong, he *should* have known?[17] But how, if he had had no exposure to any moral alternatives? Well, one might think, perhaps the relevant cognitive standards are, in such morally charged circumstances, really low:

16. We studied attributions of blameworthiness for both Susan Wolf's (1987) famous JoJo case and a case of a racist that we designed similar to the one mentioned in the text. In Wolf's case, JoJo is the son of a horrible dictator who, when he grows up, has endorsed and internalized all the values of his beloved father, and so beats peasants on a whim, and so forth. On a seven-point scale, with 7 being "most blameworthy" and 1 being "not blameworthy at all," people assign a mean score of around 6 to a control (basically JoJo's father) and a mean score of 5 to JoJo. His blameworthiness is thus viewed as mitigated relative to a seriously blameworthy control, a statistically significant difference, but not much of a difference in the real world. We found precisely the same pattern in the case of the racist we designed. Of course, our results do not settle the matter, as it has yet to be made clear whether test subjects were thinking in terms of accountability and the worthiness of *anger* in thinking about people's blameworthiness, as opposed to something like answerability, attributability, or even purely forward-looking grounds for blame (for example, to shape the character of the blamed agents in positive ways). Faraci and I discuss this lacuna in "Good Selves, True Selves" (Faraci and Shoemaker n.d.), and we hope to test subjects more carefully on precisely this dimension in future work.

17. This was our speculative hypothesis about the grounds of mitigation for moral ignorance in Faraci and Shoemaker 2014.

when it comes to what we owe to our fellows, perhaps virtually anyone ought to be able to glimpse the moral truth (even though it might be more difficult to do than it would be for someone from a nondeprived background). But if we make this move, it becomes hard to see much difference anymore between an extremely low cognitive standard for what one ought to know and there being no real standard at all; that is, it becomes hard to see any more just how we would be relying on *any* kind of knowledge condition in our blameworthiness responses. As long as one is a rational human agent, say, that alone could be enough to put one on the hook.

Indeed, this is precisely the view of many contemporary responsibility theorists, led by T. M. Scanlon (1998, 2008), Angela Smith (2005, 2008, 2012), and Matthew Talbert (2008, 2012). These philosophers deny the so-called *normative competence* condition for moral responsibility, the requirement that one must be able to recognize and respond to the moral reasons against some action or attitude in order to be morally responsible for that action or attitude. What these theorists have maintained instead is that, as long as one is capable of judging and acting on reasons *generally*, it doesn't matter if one is unable to judge or act on any countervailing moral reasons. This is why some of them say explicitly, for instance, that psychopaths may be blameworthy when they hurt us, even if they are blind to the moral reasons against doing so (for example, Talbert [2008]).

Of course, we might insist, as some do, that normative competence *is* required for blameworthiness. But we cannot dismiss out of hand the datum that many of us do and would respond with anger to the moral offenses of psychopaths. And insofar as our actual responses must at least serve as data points for theorizing—they must at least provide a rough epistemic guide to the response-independent responsibility properties—we need to take them seriously as perhaps indicative of a kind of blameworthiness for which the epistemic criterion is essentially irrelevant.

Voluntariness

The voluntariness condition has been accused of having many false positives over the years, starting with Robert Adams' (1985) paper "Involuntary Sins." Voluntariness most naturally, and perhaps exclusively, attaches to *actions*. But what Adams (1985, 4) points out is that there are plenty of morally faulty—and blameworthy—*attitudes*, including unjust anger, "jealousy, hatred, and other sorts of malice; contempt for other people, and the lack of a hearty concern for their welfare; or in more general

terms, morally objectionable states of mind, including corrupt beliefs as well as wrong desires."[18] In recent years, Angela Smith (2005, 240–50) has added to this list, counting among blameworthy attitudes our noticings (or neglectings), what occurs to us, and an even wider set of our emotional reactions, including regret, amusement, awe, admiration, and gratitude. From our initial blameworthy list, examples would include the moviegoer who is amused by scenes of domestic abuse, the employee who does everything you ask but with condescension, and the person who forgets to pick up the promised milk on the way home from work.

While some have tried, there is simply no plausible way to trace many of these blameworthy attitudes to some prior voluntary choice (Adams 1985, 4–6). This is especially obvious in cases of spontaneous emotional reactions: that the filmgoer finds the scene of domestic abuse so amusing very likely doesn't reflect any earlier voluntary choice to be amused by such things; his amusement is just a function of his warped sense of humor. Scrunch up your will as hard as you can: it is unlikely that you can make yourself be amused by graphic depictions of domestic abuse.

The argument for including such attitudes in the set of the blameworthy involves an honest recognition of the range of our responsibility responses. As A. Smith (2005, 236–37) puts it, "We quite often respond to people's spontaneous attitudes, reactions, and patterns of awareness in many of the same ways that we respond to their deliberately chosen actions—for example, with hurt feelings, gratitude, resentment, and various forms of moral and nonmoral criticism . . . , where these criticisms suggest that we do take people to be responsible and open to moral assessment for these seemingly involuntary responses." One might try to draw a theoretical, principled distinction between an "appropriate" set of responses to actions that are the result of voluntary choice and an "inappropriate" set of responses to attitudes that aren't (as does Levy 2005), but from a phenomenological, interpersonally engaged perspective, this would be arbitrary (and irrelevant). My angry response to your deliberately stepping on my foot feels identical in its blaming force and

18. Adams was writing prior to the introduction of the distinctions between types of responsibility in the contemporary literature. Perhaps, then, he would, if writing today, put his point in terms of blameworthiness of a different stripe (for example, answerability)? I doubt it. In his article, he repeatedly refers to what he's talking about as *accountability*, whereby he seems to mean by the term what most contemporary theorists mean, and he also discusses explicitly how one critical type of reproachful blame he has in mind involves *anger* (Adams 1985, 22–23).

function to my angry response to your amusement at my physically disabled daughter's struggling gait.[19]

Control

Despite Adams' and Smith's label, most of the counterexamples to the voluntariness condition in the previous section illustrated blameworthiness for the *non*voluntary, not necessarily the involuntary. Rather than being items that go against, or violate, an agent's actual will, blameworthy attitudes (like amusement at cruelty or a failure to notice something important) just aren't the sorts of things that are governable by wills generally. But if that is true, then nonvoluntary attitudes seem beyond agents' *control* as well, and so (it might be thought) they count as counterexamples to the control condition too.[20]

Nevertheless, some of those who offer counterexamples to the voluntariness condition maintain that control is a key ingredient of blameworthiness. How so? By introducing a different type of control. What governs actions is *volitional* control, they say, and it is this type of control we lack over our nonvoluntary attitudes. Nevertheless, we may still exercise what is called *rational* control over them (A. Smith 2005, 265; McKenna 2012, 194–95). The thought is that the attitudes for which we are responsible and blameworthy are governed not by our deliberate choices but by our evaluative judgments. Once we (sincerely, consciously) judge as to the worth of something, the associated emotional reactions, noticings, and so forth, ought to get in line with that judgment (and will, to the extent that we are rational). As long as an agent's attitudes are "rationally connected . . . to her underlying evaluative judgment" (A. Smith 2005, 262), those attitudes are attributable to her in a way that grounds various responsibility responses.

There are, however, false negatives even for a rational control condition on blameworthiness. First, judging as to the worth of other people isn't enough to generate various kinds of *perceptual states* constitutive of viewing them with moral regard, yet our blaming responses

19. Perhaps we would think that the amused person should have done something earlier to alter his disposition to be amused in such circumstances? This strikes me as fairly implausible. In such cases, our anger seems to be directed to the occurrent amusement, not to any previous decisions, and anyway it is entirely unclear how one might have the wherewithal to be able to tweak one's disposition to be amused in such circumstances directly.

20. There are also several examples of the nonvoluntary in Sher 2006.

range across such perceptual failures as well. Part of what is required in judging you to be valuable (and so worth regarding) is coming to see what you value from your perspective (and so seeing it as valuable in the way that you do). This is how certain facts about what you value come to seem to be reasons from within my own evaluative (practical) framework (see Shoemaker 2015, 97–103). Indeed, sometimes we explicitly demand this kind of perspectival stance of one another: "Just try to see it from my point of view!" we may say. And we surely get angry when people fail in this respect, as in, "You simply have no idea what I am going through!" But coming to see what matters to you from your perspective isn't a function of rational control. It is what I have elsewhere labeled "evaluative regard," and it is most fundamentally a perceptual state that is governed not by judgment but by *empathy* (Shoemaker 2015, 100–103).[21] A failure of evaluative regard alone—a failure to have facts about others' interests even appear to one as putative reasons—is sometimes, we think, blameworthy.

There are also some spontaneous emotional reactions that aren't subject to rational control, reactions that are instead just nonrational. Attitudes toward one's family members or attitudes of sports fans about their teams illustrate the point well. The mother of a serial killer may still well up with tears on hearing of her son's execution, despite her sincerely judging him to be a worthless human being (Shoemaker 2015, 55), just as a true fan of a perennially losing baseball team is going to get upset when the team inevitably fails to win the World Series again, even though she judges that rooting for the team just isn't worth the pain and misery it causes. These nonrational attitudes still ground plenty of responsibility responses. Neither volitional nor rational control is necessary for them.[22]

False Positives

To this point, I have focused solely on false negatives, prima facie counterexamples to each condition's purportedly being necessary to blameworthiness. But there are also three important kinds of prima facie false positives for the *set* of these conditions, cases in which knowing, voluntary,

21. So then why not call such governance a form of *control*? One can if one likes, but it is a very different form of control than rational and volitional control, which are the only forms that have been advanced in the literature as conditions on blameworthiness, so it remains a counterexample to the condition as it has been construed up until now. I am working on this issue in my unpublished paper, "Empathic Control?"

22. Perhaps these are not warranted responses, though? I deal with this possibility in discussing skepticism in a later section.

and controlled bad actions are nevertheless ones for which agents may not be blameworthy. The first kind of counterexample is offered by Michael McKenna (2012, 19):

> Consider the case of a woman whose clear moral obligation is to save a small group of people from certain death. As it happens, her child is in danger of a nontrivial harm, though not life-threatening. She cannot both save the group and aid her child, and in the absence of any "agential impediments," she decisively settles upon protecting her child. She freely does so knowing that she is violating her obligations. . . . In this sort of case I believe that we can make good sense of a person knowingly and freely doing morally wrong but yet not being blameworthy.

To make things more dramatic, McKenna appeals to Sophie's Choice, a case in which a mother has to give up one of her children to be killed by the Nazis. No matter what she does, she does wrong (knowingly and freely), but she is not blameworthy. Contrast Sophie with her evil twin Cruella, who when placed in such circumstances gives up one of her children with glee, thinking "Good riddance!" (McKenna 2012, 20). Cruella *is* blameworthy, even though she is in exactly the same tight spot as Sophie. What is the difference, and so what is the additional condition ostensibly needed for blameworthiness? *Poor quality of will*. Cruella has it; Sophie and the mother in the original case above do not.

One might be suspicious about McKenna's verdicts about blameworthiness or wrongdoing in these cases (see Shoemaker 2014 for discussion), but it still seems that *some* such cases are possible. If we grant them, one might then put forward *quality of will* as an additional condition on blameworthiness. But even if we do that, there remain false positives, namely, people with identical qualities of will (and knowledge, control, and voluntariness) whose different histories ground different responsibility responses. So of two thieves who steal out of greed, suppose one came to be the way he is via typical childhood development, whereas the other came to be the way he is via some sort of external manipulation (a nefarious neuroscientist) or incapacity (Fischer and Ravizza 1998, chap. 7; Pereboom 2014, chap. 4; Mele 2006, chaps. 6–7; Vargas 2013, chap. 9). To handle such cases, then, perhaps we should add a *history condition*: beyond having knowledge, voluntariness, control, and poor quality of will, blameworthy people must also have come to generate the bad action or attitude via some kind of acceptable historical pathway (see, for example, Mele 2009).

This addition is still not sufficient, however, to ward off false positives in the form of moral luck. So consider the final case on the blameworthy list, where two exactly similar people drive through the same neighborhood and are identically reckless in texting. The first makes it through without incident. The second hits a dog that happens to run out in front of the car. Had the dog run out in front of the first driver, that driver would also have hit the dog. Both agents, we can stipulate, knowingly, voluntarily, with the requisite amount of control, and *with identical qualities of will and histories* were recklessly texting while driving. Nevertheless, the second person will seem to many (especially to the dog's owners!) as worthy of more blame than the first, despite both drivers meeting all these "conditions" for blameworthiness to the same degree (for the original presentation of a case like this, see Nagel 1979, 28–29; for discussion along these lines, see Scanlon 2008, 125–28).

Now one response might be to deny the existence of moral luck and so refuse to view the actual degree of blameworthiness attached to the agents any differently in the two cases. There are many varieties of this move, and they all have to explain away our different reactions toward the two drivers. One might explain the differences epistemically, for example, by claiming that we are actually ignorant of the intentions of the drivers, and we tend to read off intentions from actions (Richards 1986; Rescher 1993). Or one might say that it is reasonable to demand differential responses as part of a demand for virtue, in light of (mere) causal responsibility for tragic outcomes (Wolf 2001). Or one might say that the difference in the family's reaction is explained by the greater significance of the second driver's action for them (Scanlon 2008, 159). But all of these explanations have been plausibly questioned (see Nelkin 2013 for an overview).

Regardless of which explanation one adopts, however, denying moral luck requires biting a huge bullet, as shown by Michael Zimmerman in a series of powerful articles (Zimmerman 1987, 2002, 2006, 2015). Once we deny moral luck of various sorts, we are forced, ultimately, to the position that we are responsible (and blameworthy) for tons of counterfactuals, situations in which we *would have* done something bad were it not for various circumstances or our contingent psychological profiles. The (absurd?) implications of denying moral luck are rife.

But if we accept moral luck, it seems that we must also accept that blameworthiness is not just a function of control, knowledge, voluntariness, poor quality of will, and the right history. Something even more is needed for the response-independent theorist to capture all of our

responses. But what could that be?[23] And why should we think that it will not again be threatened with false negatives with respect to its status as a necessary condition and false positives once it has been incorporated with the other conditions? Fortunately, there is a better way.

The Burden-Shifting Argument

We have seen prima facie false negatives for every response-independent proposal for the necessary conditions of blameworthiness, and we have seen prima facie false positives for every collection of sufficient conditions. It is difficult to see what response-independent natural features might impose unity on the wide array of activities and attitudes we deem blameworthy. Thus the better account—simpler, more plausible, and with greater explanatory value—is that our *emotional responses themselves* are what impose unity on the list. Just as what imposes unity on pratfalls, sly puns, silly walks, sophisticated satire, stupidity, and dead baby jokes is ultimately our fitting amusement at such things, so too what imposes unity on insults, assaults, obliviousness, insensitivity, disrespect, forgettings, negligence, omissions, dutiful "kindness," condescension, bad moral luck, and so on, under the rubric of the blameworthy is just that they are the fitting targets of anger. This is to embrace the following view:

> **Fitting Response-Dependence about the Blameworthy:** The blameworthy (in the realm of accountability) *just is* whatever merits anger (the angerworthy); that is, someone is blameworthy (and so accountable) for X if and only if, *and in virtue of the fact that,* she merits anger for X.

Just as one cannot characterize the nature of the funny without constitutive reference to our sense of humor, so too one cannot characterize the nature of blameworthiness (and so accountability) without constitutive reference to our sense of anger.

If the Anger Fits . . .

There are three items whose relationship we have been exploring: anger, the angerworthy, and the blameworthy. The blameworthy is the value we are trying to understand. The dispositional response-dependent theory ties the blameworthy to our actual anger, but this approach fails to have

23. Some even think that nonagential factors may be part of the conditions of responsibility, for example, one's moral ecology or the context in which one acts. See Sneddon 2005; Vargas 2013, 243–49; and Ciurria 2015.

the requisite normativity: only what is *worthy of*, or merits, anger could count. The question then becomes about what the relation is between the angerworthy and the blameworthy. The response-independent theorist says that the response-independent property of the blameworthy (that it was a bad action performed with voluntariness, control, knowledge, and so on) is what makes anger appropriate. The fitting response-dependent theorist says, to the contrary, that the blameworthy just *is* the angerworthy, so that one cannot make reference to the blameworthy without making reference to our fitting angry responses. What unites all the items on the blameworthy list, therefore, is simply that they are all agents toward whom anger is fitting.

The fitting response-dependent view of blameworthiness therefore obviously places all of its real weight on what the *fittingness* of anger consists in. This notion, as with the fittingness of amusement, is ambiguous. It could refer to what makes someone blameworthy for any individual action or attitude, or it could refer to what makes those blameworthiness-makers count as such. Consider an example: I stand you up for a lunch date without telling you because I really just felt like staying home to watch a rerun on TV. Once you find out, you get angry with me. What makes your anger fitting? For ease of discussion, let us suppose that it was primarily my poor quality of will, that is, your anger appraises my quality of will as poor, and so your appraisal is correct (fitting) just insofar as my quality of will *is* poor. Poor quality of will is an objective agential feature of mine, consisting, roughly, in what my various reasons, motivations, cares, and concerns are, as well as perhaps their role and weight relative to one another in my deliberations (see, for example, McKenna 2012, chap. 3). If I didn't in fact have poor quality of will, it would be unfitting for you to respond as you did. But if my objectively poor quality of will is what makes your anger fitting, then it may seem as if blameworthiness isn't a response-dependent matter after all.[24] Rather, my blameworthiness seems to be a function of *my poor quality of will*, and so I merit anger only in virtue of being antecedently (and response-independently) blameworthy.

Analogously to the funny, though, the fundamental fitting response-dependent feature of the theory is really about what makes certain objective features the *anger fitmakers* in the first place. So to the extent that control, knowledge, voluntariness, quality of will, or history are indeed among those fitmakers, they are so *in virtue of* their triggering

24. Or it is only partially such. See McKenna 2012, chap. 2.

our anger sensibilities. They are just the sorts of properties to which we humans are built to respond with a heated demand for acknowledgment or a tendency to retaliate. There is no better way to explain the motley collection of blameworthy fitmakers otherwise (goes my burden-shifting argument). Once we really think about the weird and wide variety of things that we tend to respond to with anger, it becomes more and more difficult to maintain the view that there is some unified set of response-independent responsibility properties behind them all. Instead, their unity is just a function of this particular sort of heated response.

The story is even more compelling when we think again about all the prima facie false positives to response-independent attempts, where meeting every single one of the long disjunction of proposed response-independent conditions may still be insufficient to explain apt variations in type and degree of anger (as in the moral luck cases). What the fitting response-dependent theorist can easily say here is that the variations are all explained by factors that, in circumstances like these, *tend to produce corresponding variations in fitting anger*. For instance, mere causal responsibility alone ordinarily produces no such variations. If I accidentally bump into you and cause you to spill your milk, I merit no more anger for that than if the wind did it. But with the two drivers, the second driver's causal responsibility for killing the dog is the only difference between the two cases, and to the extent that there seems to us to be a difference in how much anger each merits, that single difference has to be its source. But then why does mere causal responsibility matter in *this* case when it doesn't matter in others? The most plausible explanation is that it is an instance of causal responsibility *in the way that makes us angrier*, in the way that motivates us to more heatedly tell the second driver off.

As with the amusing, there may be serious disagreement about the angerworthy. The interesting sort of disagreement would be about which properties our human sensibilities have rendered the fitting anger-triggers (that is, the blameworthy-makers). Now, much of the time our "disagreement" will actually have its source in various obscuring factors (D'Arms and Jacobson 2010): we may be too tired or depressed to feel anger at what merits it, or it may be too hard for us to see that certain fitting-anger-makers are actually in place. But let us suppose that we have cut through the obscuring factors and we still disagree over what makes some feature count as angerworthy. Perhaps, say, we disagree over whether a failure to notice a romantic partner's new hairstyle is angerworthy. In trying to resolve this dispute, won't we have to appeal to response-inde-

pendent grounds for why one of our views has normative authority over the other?[25]

We will not. To see why, it behooves us to pay attention to the way in which such disputes actually occur. When we seem to be at loggerheads, we often ask each other, "What makes you a good judge in these matters? Why should I defer to your normative authority?" Here we are asking the other person to defend the refinement and development of his or her anger sensibility (see, by way of comparison, Gibbard 1990, 191–92). This is because we know that there are defective senses of anger that are sometimes the source of disagreement, where incorrect angry responses are generated by dysfunctional human machinery, the product perhaps of coddled or brutalized youth. So in asking for a defense of sensibilities, we are asking for reasons to trust another's judgment over our own. Now these reasons may themselves be, yes, response-independent, consisting in objective facts about what makes some sensibilities better—more sensitive and attuned to the value in question—than others. But then once we have determined that one sensibility is indeed better than another in these matters, the properties that that sensibility identifies as blameworthiness-makers will of course be response-dependent.[26]

Suppose, though, that even after all this we remain at loggerheads. At this point, we can typically only urge each other to look harder or to try and see it in the way that we or a range of others do, to tell a coherent epistemic story tracing a recognizably human route to our verdicts, or to prod each other's sense of embarrassment when lacking certain answers (see Gibbard 1990, 192–94). But even these procedures may leave our dispute unresolved. Fundamental disagreement is thus possible, as at the end of the day our equally refined sensibilities may just crank out different responses to the same thing, and there are no response-independent factors in such cases to which we can point to explain why some properties count as blameworthy-makers and some do not.

Thus ends my exposition and burden-shifting argument in favor of the fitting response-dependent view. As far as I can determine, there are three skeptical worries one might have about it. To bolster and fully

25. My thanks to an anonymous referee for pressing me to address this issue explicitly.

26. Compare to a sommelier, whose wine sensibility is presumably more refined, objectively, than that of the diner, but who still provides response-dependent (that is, taste-based) reasons for choosing one wine over another.

defend the response-dependent view, therefore, I need next to articulate and defend against these worries.

Objection 1: Response-Drenched (Meta)skepticism

The first skeptical worry, what Sommers (2012) calls "metaskepticism," is that I have improperly downplayed the scope and implications of fundamental disagreement. Sommers, for instance, points to widespread disagreement in our responsibility responses, and then infers that there just is no universal responsibility sensibility from which "the" response-dependent understanding of responsibility could even be drawn. Throughout various historical epochs or disparate contemporary cultures, he notes, there are plenty of examples of people viewing agents as blameworthy that are in sharp contrast with how contemporary Westerns would view them. In ancient Greece, some faultless agents (for example, Oedipus and Agamemnon) were treated as blameworthy (Williams 1993, 68–72, 132–35; Sommers 2012, 35). In honor cultures, punishment may be carried out against siblings or fellow tribemates of the guilty party (Sommers 2012, 34; Boehm 1984). And there are plenty of alternative contemporary morality systems that involve moral norms and blame for violations thereof that many of us might find bizarre, for example, failures to live up to personal ideals (Williams 1993, chap. 4; Russell 2013), or betrayals of the values of sanctity or purity (Graham, Haidt, and Nosek 2009; Haidt 2012). If all we really have to theorize about, therefore, are our various quite local responses, and these are highly variable, then, according to Sommers (2012, 108, "There are no universally true conditions for moral responsibility. . . . Since considered intuitions ultimately ground our theories of responsibility, there does not seem to be a principled way of establishing conditions of moral responsibility that would apply across cultures." Bernard Williams's (1993, 55) version of this worry is less sweeping but no less confident: "There is not, and there never could be, . . . just one correct conception of responsibility." Call this *response-drenched skepticism*.

Nevertheless, a fitting response-dependent theory grounded in universal human sentiments is perfectly compatible with substantial amounts of cultural variation in the precise objects to which those sentiments respond. This would be true if, for example, there were several such objects, all of which tend to trigger (properly developed and informed) human sensibilities, just to different *degrees* in different cultures, perhaps depending on ecology and social structure. And as I have

explicitly noted earlier, there may be other appropriate blaming emotions in addition to anger, including disdain, disapproval, and contempt. Some cultures may emphasize some of these blaming emotions over others—perhaps indeed due to ecology and social structure—and when they do, they are likely to have responsibility responses of a different overall hue than those who emphasize other emotions. So a more aretaically minded culture, like that of the ancient Greeks, may emphasize character (as targeted by emotional responses like admiration, disdain, and shame) over disregard (as targeted by anger). This may also be true of "purity" cultures. But the fact (if it is a fact) that the ancient Greeks put most emphasis on aretaic emotions in their responsibility exchanges doesn't at all mean that they too didn't feel anger in many ways just as we do and toward the same objects as we do. Indeed, Aristotle's characterization in *Rhetoric* of anger as fitting slights and motivating "conspicuous revenge" still powerfully resonates with us. So their different emphases do not imply that they had different anger sensibilities from us or that we haven't already achieved significant convergence on the multitude of blameworthy fitmakers our anger sensibilities have carved out.

It is also important to note that sometimes what makes other cultures' blameworthiness practices feel foreign is the form the *expression* of their anger takes. But how one communicates one's anger is subject to ethical norms, not fittingness norms about anger as such, and there is nothing in the fitting response-dependent view that predicts no cultural variation in the former. Consequently, the response-drenched worry may be set aside.

Objection 2: Theory-Drenched Skepticism

The second skeptical response says that an alternative option when a response-independent theory of the blameworthy is faced with prima facie counterexamples is to bite the bullet. Once the conditions for blameworthiness have been independently and antecedently established, so much the worse for angry responses that conflict with the theory's predictions. Indeed, goes the argument, we need to eliminate or revise a lot of our actual blaming responses, given how few of our actions or attitudes meet the true theoretical conditions for responsibility. Call this the *theory-drenched* skeptical approach. It is found in the work of Gideon Rosen (2004), Michael Zimmerman (1997, 2002), Neil Levy (2011), and Derk Pereboom (2014), among others.

Consider Gideon Rosen's version. He starts with the widely accepted datum that ignorance excuses from blameworthiness, but then he notes that it does so only if such ignorance is nonculpable. But as it turns out (for a variety of reasons), we cannot exclude the possibility that *most* people's ignorance is nonculpable. Consequently, we have insufficient grounds to judge in any case that others are responsible for their actions, and so we "should presumably stop blaming them" (Rosen 2004, 311). And what Neil Levy does is similar with respect to the control condition. Luck excludes control. But, he argues, luck is thoroughgoing in our lives (from the formation of our traits to the circumstances in which we find ourselves), and so we have no good reason to believe that anyone can really control, and so be truly blameworthy for, his or her actions (Levy 2011).

I have two connected replies to these theory-drenched skeptics. The first is to run the old "Modus Tollens/Modus Ponens" move, that is, to point out that what is an embraced implication to one person is a reductio to another. The reductio here involves reminding these theorists (and others inclined toward such skepticism) of the significant cost of the theory-drenched approach, which is that it divorces theory from much of our humanity in a particularly stark way. In particular, it disconnects our emotions from the values they are purportedly tracking (D'Arms and Jacobson 2010, 610–11), and it does so on highly contestable theoretical grounds. This is, of course, Strawson's point, but it bears repeating: to demand thoroughgoing revision to our system of reactive attitudes is to demand what is well-nigh psychologically impossible, but even if it is not, it could only be effected at the cost of interpersonal relationships, which are essentially a matter of being vulnerable to such reactions (Strawson 2003, Shabo 2012). While Strawson's claims here are very probably too strong (see, for example, Wallace 2014), the point remains that interpersonal responsibility practices—as opposed to institutional legal responsibility practices[27]—are emotionally drenched. Indeed, Strawson claims, and I agree, that our responsibility practices—how to treat or sanction those who are on responsibility's "hook"—are built on *top* of the reactive attitudes, as partial expressions of them (Strawson 2003, 92). The emotional responses are the foundational core of responsibility, and they are natural appraisals whose appropriate triggering conditions are deeply instilled in us as a function of ordinary human development and learn-

27. Which, I should note, Rosen (2004, 296) (and many others) explicitly analogizes to moral responsibility.

ing. Counseling against the emotional responses themselves via theoretical metaphysical arguments would be like counseling against envy via Stoic arguments or counseling against fear of death via Buddhist arguments: it might successfully effect emotional changes in some, but these are going to be people with almost superhuman capacities.[28] For the rest of us, these theoretical arguments will have a hard time gaining any grip on our emotional lives.

But won't this mean our responses *are* inconsistent? If so, how can we as philosophers let this inconsistency stand! This is an unfounded worry, for a fitting response-dependent theorist can allow for responses with fine-grained fittingness conditions that can defuse the purported inconsistency. Anger, for example, could be appropriate for *this* type of luck but not *that*. Or it might appraise some types of epistemic gaps as culpable and some others as not. It would thus not be luck or epistemic gaps per se that ground excuses, and so the "worrisome" extension from some cases to all (or almost all) could not be made. Further, even if our responses are inconsistent in some sense, the fitting response-dependent theorist can have a nonchalant attitude about it. After all, we seem to find some kinds of nonstandard walks funny (think of the Ministry of Silly Walks) but find other kinds of nonstandard walks not funny (think of slight limps, long strides, and military marches). There seems to be no answer as to why some are funny and some aren't, except that that is just the way we have been built to respond. So too the objection of inconsistency will have far less force against the fitting response-dependent theorist than against the response-independent theorist. The fact that we "inconsistently" find some cases of luck or ignorance excusing and others not may thus not be, on its own, a reason to adopt a revisionary line against any of our responses.

There is a second general response to the theory-drenched skeptic, a "hoist with your own petard" reply. Recall that the skeptical strategy starts with some allegedly uncontroversial datum about excuses ("ignorance/luck excuses from blameworthiness") and then shows how the excusing condition is far more insidious than we might have thought, requiring pervasive response revision on our part. But what is the source of the original datum? *Our responses*. Specifically, we are given "uncontroversial" cases and asked to respond to them. So, for example, what would

28. Thanks to Chandra Sripada for discussion. Even Buddhist masters self-report that their fear of death is extremely high, despite their theoretical metaphysical commitments that death should not be feared (Garfield et al. 2015).

we think if Jekyll had put poisoned sugar in Hyde's tea, but Jekyll had no idea it was poisoned and was not culpable for his ignorance? (Rosen 2004, 299–300) The question is then put: Is Jekyll blameworthy for what he did? Clearly, the answer is no. But how do we determine this clear answer? By consulting our *feelings*, surely: Would we be angry at Jekyll? I see no other way to arrive at a verdict in these cases (except by induction, perhaps, from previous cases in which we have already consulted our feelings). We are only then led down the garden path to seeing why most of our *other* responsibility responses need revision. But on what basis can the theory-drenched skeptic privilege the motivating emotional responses over all the to-be-revised rest of them? Could it be the clarity of the response? Hardly, as the gut punch of theory-drenched skepticism partly consists in just how many other "clear" responses in fact have to be revised. I am hard-pressed to think of any other basis for privileging one set of responses over the other. The revisionary move seems blocked from the get-go.[29]

Objection 3: Shadow Skepticism

The third response is *shadow skepticism*, according to which the angerworthy isn't what, on its own, matters; rather, it shadows some other response-independent property constituting what we truly value instead. Anger might thus be a rough-and-ready guide to the blameworthy, but it is irrelevant to its actual constitution qua value (see, by way of comparison, Arpaly 2006, 28, 31).

This is, of course, response-independence redux. So one might run a version of shadow skepticism that appeals to response-independent properties like the ones already surveyed, having to do with voluntariness, control, knowledge, quality of will, and so forth. I won't rehash my arguments against those options (although some of them will be revisited briefly below). Instead, I want to consider a different angle into shadow skepticism, what several naturalistically inclined theorists have advanced as the most plausible story about anger's proper target, and then use that

29. This argument is akin to a rejection of Peter Singer/Joshua Greene-style "evolutionary debunking arguments" about nonutilitarian ethics, an argument that points to their being grounded on dubitable emotional intuitions. But, of course, utilitarian arguments are grounded on plenty of the same kind of "dubitable" emotional intuitions (for example, intuitions about saving a girl in a drowning pond, or general intuitions about responding to suffering), so it is hard to see any reason to privilege their emotional intuitions over nonutilitarian ones. See Singer 2005; Greene 2008; and the source of this objection in Kahane 2011.

account of anger's function to tell a story about the response-independent property anger might be thought to fit. The story would go like this:

> Anger evolved as a tool for enforcing social norms. Indeed, this is the only way social cooperation could have emerged, with its norms policed by psychological mechanisms recruited for doing so. Our angry passions commit us in the right sorts of ways to this task, serving the broader rational goals via actual enforcement of norm violations, but also serving as credible threats and deterrents to others (so anger has both an action-priming and a signaling function).[30]
>
> Given, then, that anger evolved in this way to response-independent features of our environment (that is, to violations of social norms), its function must be just that, to enforce social norms. Of course, not just *any* norm violations are to be enforced, as some norm violations are benign (and so sometimes render amusement fitting instead). *Being a nonbenign norm violator*, therefore, is the property that people must actually be interested in when it comes to anger; indeed, what reason would we have to care about the angerworthy unless it were just identified with the response-independent property of being a nonbenign norm violator? This is the property constituting the blameworthy, then, and while anger might still be our best epistemic marker for identifying these blameworthy violators, it should bear no metaphysical weight in the construction of blameworthiness itself.[31]

I have two responses to the shadow skeptic. The first challenges her to characterize the property of being a *nonbenign norm violator* in response-independent terms. I will argue, to the contrary, that this cannot be done. Instead, a plausible account of what nonbenign norm violations and violators are, and why they matter to us, can be made only by covertly appealing to anger.

Consider a case in which a part of my body knocks into yours, causing an injury. If I am pushed into you, I commit no nonbenign norm violation, because *I* don't do anything. If, with a gun to my head, I punch you, I don't do anything *voluntarily*. If I am blind and stumble

30. This is a story told by many, including Frank (1988), Nichols (2002, 2007), Nichols and Mallon (2006), Henrich and Henrich (2007), Boyd and Richerson (2005), and Sripada and Stich (2007). In its presentation, I have followed the nice summary presented by McGeer (2013, 171–72).

31. Again, I'm indebted to the February 2014 discussion on PEA Soup (blog), where a form of this objection, as articulated by Schroeter (2006), was summarized and discussed by Justin D'Arms and Daniel Jacobson with respect to some other values, http://peasoup.typepad.com/peasoup/2014/02/featured-philosophers-darms-and-jacobson.html.

onto your foot, then while it's true now that I voluntarily do something (in walking), I nevertheless don't walk into you *intentionally*. If my knocking you down is the only way to save an infant about to fall into a pool, it doesn't seem that I violate a *norm*, given that norms against running into others surely have implicit emergency "out" clauses. So nonbenign norm violators would seem to be agents who voluntarily and intentionally break the demands or expectations of some norm in its fully spelled-out, nuanced form (where this includes an explicit rendering of its implicit "out" clauses). Will this do as the requisite response-independent analysis of the blameworthy?

Not even close. There are those agents, for instance, whose blameworthy doings are *not* intentional, as when in recklessly skateboarding I run into you and knock you down, or as when I call you excitedly and start talking about the huge soccer upset that just occurred, a game you had recorded to watch in blissful ignorance later. In neither of these cases (nor in many, many others) do I intend to violate any norm, yet I may nevertheless be blameworthy. And the blameworthy also isn't restricted to things that an agent voluntarily *does*. Recall that some omissions are blameworthy, as are many attitudes.

Finally, what precisely shapes or gives rise to the implicit emergency "out" clauses of various norms, making some performances of their stated proscriptions blameworthy and some not? Suppose, for instance, that instead of stepping on your foot to save an infant, I do so while amusing myself by counting the steps it takes to get from one wall to the other? That would merit anger. But what could the relevant (response-independent) threshold between blameworthy and nonblameworthy exceptions to the norm consist in? It is entirely mysterious.[32]

Recall the shadow skeptic's basic argument: any reason we have to care about what anger fittingly appraises is just parasitic on what we *actually* have reason to care about, namely, the response-independent property that anger was naturally selected to police: nonbenign norm violations. While anger might be a good epistemic tool for tracking this property, it cannot be a metaphysical constituent of it. I have responded by showing just how difficult it is to vindicate this claim, as we cannot get

32. One might try to define the threshold by leaning on the distinction between the moral and the nonmoral, but this won't do, as we also lack a plausible response-independent distinction between these two domains. See Sinnott-Armstrong and Wheatley 2012 and 2014, as well as arguments against the so-called "moral/conventional" distinction in Kelly et al. 2007, and Shoemaker 2011a.

any clear, counterexample-free characterizations of "nonbenign" or "violations."

So how *do* we have a (roughly) unified vision of the actions and attitudes falling under the blameworthy rubric; that is, how is it that some of those who intentionally and voluntarily slight or hurt others, say, are grouped together as "nonbenign norm violators" with agents who unintentionally or involuntarily slight or hurt others? And how do we determine what our norms' "out" clauses are in a way that enables us to determine when those norms are not in fact being violated? Again, the most plausible answer to these questions seems to be that what count as nonbenign norm violators are (all and only) agents whose actions or attitudes tend to strike us *in the angry way*.[33] This is the first response to the shadow skeptic.

Here is a second response. The anger that we have in response to certain actions and attitudes is a function of what we care about, which is indeed a function of our specific biological and cultural history, our *anthropological* nature. The shadow skeptic's claim, though, seems to be that we evolved to care about norm violators and so evolved to get angry with them *because* doing so promoted our ancestors' reproductive success. That is to say, there was a (response-independent) property—nonbenign norm violator—that some agents responded to with anger, where

33. Why not say, as both Michael McKenna and Nathan Stout have (in private correspondence), that "nonbenign norm violators" are all unified under the rubric of "agents who violate our demands for a reasonable degree of good will"? I think this is probably the best *gloss* on what anger fits, but the fact that we can characterize that domain with a snappy formula like this does not at all mean that it marks some response-independent property that fitting anger shadows. We can see this by running the same moves on this formulation: What counts as a "failure of a reasonable degree of good will"? The answer will be so context-, relationship-, and person-sensitive, I believe, as to be adequately captured only by a massively disjunctive analysis that is more plausibly characterized, again, as "those actions or attitudes that strike us in the angry way." For just one example, suppose that every year George buys his wife Martha the same ugly carnations that she hates for their anniversary. Martha desperately wishes he would do something surprising for her, but she can't tell him that, as then it wouldn't be a surprise. So this year he once again gives her carnations and she gets angry at him (Shoemaker 2011b, 620; see Smith 2012 for comments on the case that I now find more plausible). For nearly anyone else, George's "means-well" gift giving would count as being of sufficiently good will that he would be off the anger hook. But if meaning well in giving a gift to a loved one doesn't count as a "reasonable degree of good will," what does? Yet in this case Martha's anger seems fitting. So we can add a disjunct to our list of what counts as "reasonable degree of good will" to include this case—a dashing of hopes, say—but where does it end? Just call it what it is, a very rough gloss on the kinds of things that strike us in the angry way, and call it a day.

doing so somehow increased their fitness, so that we, their descendants, inherited that response mechanism. But this would be quite incredible, if true. Rather, what seems more in line with the science on this score is that our early ancestors responded with anger to some people because doing so made their targets behave in certain ways and made the angry people themselves behave in certain ways (so as to avoid the anger of others or provide some prudential reputational benefit [Frank 1988, 110–11]), and such behavior promoted their reproductive success, so was inherited by us.[34] On this latter explanation, there is no tracking of response-independent properties being done at all. Indeed, bolstering this explanation is the fact that it is entirely unclear just how the (response-independent) fact of something's being a nonbenign norm violation per se could make *anger* at the violator reproductively advantageous.

The only other explanation the shadow skeptic could give, then, would be that evolutionary forces have either pushed our angry responses away from, or rendered them at best random with respect to, the response-independent "nonbenign norm violator" properties. But then it would be a sheer coincidence if our angry responses *ever* hit on the correct properties of being a nonbenign norm violator.[35] Remember, the shadow skeptic's thought is that the property of being a "nonbenign norm violator" is what *really* interests us, yet fitting anger is still a good epistemic indicator of that property. But what possible reason could we have to rely on anger's epistemic tracking powers if its evolutionary selection was at best random with respect to that property? And if anger is no better than chance at tracking that property, then there is no reason to think that what it appraises *shadows* that property at all.

Consequently, there are very good reasons to resist the shadow skeptic, about both the funny and the blameworthy, and these responses run on parallel tracks. The most plausible story about both is a response-dependent one.[36]

34. For versions of an argument like this in metaethics with respect to normative judgments, see Street 2006, 2008, 2009, and 2011. For sharp critical discussion of this argument, see Berker 2014.

35. Again, this general move is drawn from Street 2006, 121–25; 2008, 208–9; 2009, 233–34; 2011, 13–14; and discussion in Berker 2014, 223–24.

36. A complete defense of the response-dependent view would need to show it to be better than McKenna's (2012, chap. 2) compromise proposal, what he calls the modest metaphysical view. According to him, holding responsible metaphysically depends on the facts about being responsible, but being responsible *also* metaphysically depends on the facts about holding responsible. McKenna renders this circular view nonvicious by claim-

Conclusion

I have said way too much already, but I will briefly mention two implications response-dependent responsibility has for our theorizing about responsibility generally, one corrective, one predictive. First, if I am right, then investigating the nature of the blameworthy (in the accountability domain) reduces to a matter of investigating the fittingness conditions of anger. Some theorists already engage in this general method, but often it is done only as a way of revealing what response-independent properties our emotional responses allegedly track. On their approach, it is an open question whether our set of responses might need or lack independent justification, and so an open question whether, for instance, determinism's truth could undermine responsibility. On my approach, asking for an external (response-independent) justification of the angerworthy would, as Strawson (2003, 83, 93) said, miss the point, for it would be to ask for a justification for being human.[37]

The second implication has to do with future directions of research. I have made the case for fitting response-dependence here only with respect to anger and accountability-blameworthiness. But as I have noted along the way, there are many other emotional responsibility responses, including admiration, disdain, contempt, regret, pride, and gratitude. A thorough defense of response-dependence about responsibility across the board would have to say something about all of these as well. I am currently unsure what to say about some of them, but my hunch is that when the responsibility responses in question are human sentiments (pan-cultural, irruptive, emotional appraising dispositions, defined and differentiated by their action tendencies), they are going to (fittingly) aim at properties that will probably be quite different from the properties picked out by fitting anger. In other words, different responsibility emotions could target very different agential capacities, which could well mean that they implicate different types of responsibility (Shoemaker 2015, part I). Nevertheless, it is possible as well that only

ing that holding responsible must answer to certain prior agential facts, but those agential facts are in part about a facility in holding responsible. I simply lack the space to address McKenna's nuanced and interesting view adequately here. In brief, though, I believe that what makes something count as an agential feature in the first place (as opposed to a merely nonagential psychological feature like the ability to perceive colors, suffer, or recognize faces) is itself fundamentally a function of our fitting response-dependent commitments.

37. I owe this phrase to Sean Foran.

some of these alternative responsibility responses have response-dependent targets, which means the remainder could be mere epistemic trackers of response-independent targets. This would make for a very messy enterprise, but I do not think there is anything simple or uncomplicated about human beings and their emotions generally, so this result would not surprise me in the least.

References

Adams, Robert Merrihew. 1985. "Involuntary Sins." *Philosophical Review* 94: 3–31.

Arpaly, Nomy. 2006. *Merit, Meaning, and Human Bondage*. Princeton: Princeton University Press.

Bennett, Jonathan. 1980. "Accountability." In *Philosophical Subjects*, ed. Zak van Straaten, 14–47. Oxford: Clarendon Press.

Berker, Selim. 2014. "Does Evolutionary Psychology Show That Normativity Is Mind-Dependent?" In *Moral Psychology and Human Agency*, ed. Justin D'Arms and Daniel Jacobson, 215–52. Oxford: Oxford University Press.

Boehm, C. 1984. *Blood Revenge: The Anthropology of Feuding in Montenegro and Other Tribal Societies*. Lawrence: University Press of Kansas.

Boyd, Brian. 2004. "Laughter and Literature: A Play Theory of Humor." *Philosophy and Literature* 28: 1–22.

Boyd, Robert, and Peter J. Richerson. 2005. *The Origin and Evolution of Cultures*. New York: Oxford University Press.

Brink, David, and Dana Nelkin. 2013. "Fairness and the Architecture of Responsibility." *Oxford Studies in Agency and Responsibility* 1: 284–313.

Bykvist, Krister. 2009. "No Good Fit: Why the Fitting Attitude Analysis of Value Fails." *Mind* 118: 1–30.

Ciurria, Michelle. 2015. "Moral Responsibility Ain't Just in the Head." *Journal of the American Philosophical Association* 1: 601–16.

Clark, Michael. 1987. "Humor and Incongruity." In *The Philosophy of Laughter and Humor*, ed. John Morreall, 139–55. Albany: State University of New York Press.

Clarke, Randolph. 2014. *Omissions: Agency, Metaphysics, and Responsibility*. Oxford: Oxford University Press.

D'Arms, Justin, and Daniel Jacobson. 2000. "The Moralistic Fallacy: On the 'Appropriateness' of the Emotions." *Philosophy and Phenomenological Research* 61: 65–90.

———. 2003. "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)." In *Philosophy and the Emotions*, ed. A. Hatzimoysis, 127–46. Cambridge: Cambridge University Press.

———. 2006. "Sensibility Theory and Projectivism." In *The Oxford Handbook of Ethical Theory*, ed. David Copp, 186–218. Oxford: Oxford University Press.

———. 2010. "Demystifying Sensibilities: Sentimental Values and the Instability of Affect." In *The Oxford Handbook of Philosophy of Emotion*, ed. Peter Goldie, 585–613. Oxford: Oxford University Press.
———. Forthcoming. *Rational Sentimentalism*. Oxford: Oxford University Press.
Deigh, John. 2011. "Reactive Attitudes Revisited." In *Morality and the Emotions*, ed. Carla Bagnoli, 197–216. Oxford: Oxford University Press.
Faraci, David, and David Shoemaker. 2010. "Insanity, Deep Selves, and Moral Responsibility." *Review of Philosophy and Psychology* 1: 319–32.
———. 2014. "Huck vs. JoJo: Moral Ignorance and the (A)Symmetry of Praise and Blame." *Oxford Studies in Experimental Philosophy* 1: 7–27.
———. n.d. "Good Selves, True Selves." Unpublished manuscript.
Fischer, John Martin, and Mark Ravizza, eds. 1993. *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press.
———. 1998. *Responsibility and Control*. Cambridge: Cambridge University Press.
FitzPatrick, William J. 2008. "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge." *Ethics* 118: 589–613.
Frank, Robert H. 1988. *Passions within Reason: The Strategic Role of the Emotions*. New York: Norton.
Frankfurt, Harry. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.
Frijda, Nico. 1986. *The Emotions*. Cambridge: Cambridge University Press.
———. 2007. *The Laws of Emotion*. Lawrence Erlbaum.
Garfield, Jay L., et al. 2015. "Ego, Egoism and the Impact of Religion on Ethical Experience: What a Paradoxical Consequence of Buddhist Culture Tells Us about Moral Psychology." *Journal of Ethics* 19: 293–304.
Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. 2009. "Liberals and Conservatives Rely on Different Sets of Moral Foundations." *Journal of Personality and Social Psychology* 96: 1029–46.
Greene, Joshua. 2008. "The Secret Joke of Kant's Soul." In *Moral Psychology: The Neuroscience of Morality*. Vol. 3, ed. Walter Sinnott-Armstrong, 35–79. Cambridge, MA: MIT Press.
Haidt, Jonathan. 2012. *The Righteous Mind*. New York: Vintage.
Hart, H. L. A. 2008. *Punishment and Responsibility: Essays in the Philosophy of Law* 2nd ed. Oxford: Oxford University Press.
Henrich, Joseph, and Natalie Henrich. 2007. *Why Humans Cooperate: A Cultural and Evolutionary Perspective*. New York: Oxford University Press.
Jacobson, Daniel. 2011. "Fitting Attitude Theories of Value." In *Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), ed. Edward N. Zalta. https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/.
Kahane, Guy. 2011. "Evolutionary Debunking Arguments." *Nous* 45: 103–25.

Kelly, Daniel, et al. 2007. "Harm, Affect, and the Moral/Conventional Distinction." *Mind and Language* 22: 117–31.

Levy, Neil. 2005. "The Good, the Bad and the Blameworthy." *Journal of Ethics and Social Philosophy* 1. http://lawjesp.law.usc.edu/PDF/a645_Levy-vol-1-no-2.pdf.

———. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.

McGeer, Victoria. 2013. "Civilizing Blame." In *Blame: Its Nature and Norms*, ed. D. Justin Coates and Neal Tognazzini, 162–88. Oxford: Oxford University Press.

McGraw, A. Peter, and Caleb Warren. 2010. "Benign Violations: Making Immoral Behavior Funny." *Psychological Science* 21: 1141–49.

McGraw, A. Peter, Lawrence E. Williams, and Caleb Warren. 2013. "The Rise and Fall of Humor: Psychological Distance Modulates Humorous Responses to Tragedy." *Social Psychological and Personality Science*. doi:10.1177/1948550613515006.

McKenna, Michael. 2006. "Collective Responsibility and the Agent Meaning Theory." *Midwest Studies in Philosophy* 30: 16–34.

———. 2012. *Conversation and Responsibility*. New York: Oxford University Press.

Mele, Alfred R. 2006. *Free Will and Luck*. Oxford: Oxford University Press.

———. 2009. "Moral Responsibility and Agents' Histories." *Philosophical Studies* 142: 161–81.

Nagel, Thomas. 1979. "Moral Luck." In *Mortal Questions*. Cambridge: Cambridge University Press.

Nelkin, Dana K. 2013. "Moral Luck." In *Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), ed. Edward N. Zalta. http://plato.stanford.edu/archives/win2013/entries/moral-luck.

Nichols, Shaun. 2002. "Norms with Feeling: Towards a Psychological Account of Moral Judgment." *Cognition* 84: 221–36.

———. 2007. "After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes." *Philosophical Perspectives* 21: 405–28.

Nichols, Shaun, and Ron Mallon. 2006. "Moral Dilemmas and Moral Rules." *Cognition* 100: 530–42.

Oring, Elliott. 2003. *Engaging Humor*. Urbana: University of Illinois Press.

Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

———. 2017. "Responsibility, Regret, and Protest." *Oxford Studies in Agency and Responsibility* 4. Oxford: Oxford University Press.

Pinker, Steven. 1997. *How the Mind Works*. New York: Norton.

Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. 2004. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114: 391–423.

Rescher, Nicholas. 1993. "Moral Luck." In *Moral Luck*, ed. D. Statman, 141–66. Albany: State University of New York Press.

Richards, N. 1986. "Luck and Desert." *Mind* 65: 198–209.
Rosen, Gideon. 2003. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103: 61–84.
———. 2004. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18: 295–313.
———. 2008. "Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility." *Journal of Philosophy* 105: 591–610.
Russell, Paul. 2013. "Responsibility, Naturalism, and 'The Morality System'." *Oxford Studies in Agency and Responsibility* 1: 184–204.
Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
———. 2008. *Moral Dimensions*. Cambridge, MA: Belknap Press of Harvard University Press.
Scarantino, Andrea. 2014. "The Motivational Theory of Emotions." In *Moral Psychology and Human Agency*, ed. Justin D'Arms and Daniel Jacobson, 156–85. Oxford: Oxford University Press.
Schroeter, Francois. 2006. "The Limits of Sentimentalism." *Ethics* 116: 337–61.
Scruton, Roger. 1987. "Laughter." In *The Philosophy of Laughter and Humor*, ed. John Morreall, 150–72. Albany: State University of New York Press.
Shabo, Seth. 2012. "Where Love and Resentment Meet: Strawson's Intrapersonal Defense of Compatibilism." *Philosophical Review* 121: 95–124.
Sher, George. 2006. "Out of Control." *Ethics* 116: 285–301.
———. 2009. *Who Knew? Responsibility without Awareness*. New York: Oxford University Press.
Shoemaker, David. 2011a. "Psychopathy, Responsibility, and the Moral/Conventional Distinction." In supplement, *Southern Journal of Philosophy* 49: 99–124.
———. 2011b. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121: 602–32.
———. 2014. "McKenna's Quality of Will." *Criminal Law and Philosophy* 9: 695–708.
———. 2015. *Responsibility from the Margins*. Oxford: Oxford University Press.
———. Forthcoming. "You Oughta Know! Defending Angry Blame." In *The Moral Psychology of Anger*, ed. Myisha Cherry and Owen Flanagan, New York: Rowman and Littlefield.
Singer, Peter. 2005. "Ethics and Intuitions." *Journal of Ethics* 9: 331–52.
Sinnott-Armstrong, Walter, and Thalia Wheatley. 2012. "The Disunity of Morality and Why It Matters to Philosophy." *Monist* 95: 355–77.
———. 2014. "Are Moral Judgments Unified?" *Philosophical Psychology* 27: 451–74.
Smith, Angela M. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115: 236–71.
———. 2008. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138: 367–92.

———. 2012. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122: 575–89.
Smith, Holly. 1983. "Culpable Ignorance." *Philosophical Review* 92: 543–71.
———. 2011. "Non-Tracing Cases of Culpable Ignorance." *Criminal Law and Philosophy* 5: 115–46.
Sneddon, Andrew. 2005. "Moral Responsibility: The Difference of Strawson, and the Difference It Should Make." *Ethical Theory and Moral Practice* 8: 239–64.
Sommers, Tamler. 2012. *Relative Justice*. Princeton: Princeton University Press.
Spinka, Marek, Ruth C. Newberry, and Marc Bekoff. 2001. "Mammalian Play: Training for the Unexpected." *Quarterly Journal of Biology* 76: 141–68.
Sripada, Chandra, and Stephen Stich. 2007. "A Framework for the Psychology of Norms." In *Culture and Cognition*, vol. 2 of *The Innate Mind*, ed. P. Carruthers, S. Laurence, and S. Stich, 280–301. Oxford: Oxford University Press.
Strawson, P. F. 2003. "Freedom and Resentment." In *Free Will*, ed. Gary Watson, 72–93. 2nd ed. Oxford: Oxford University Press.
Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109–66.
———. 2008. "Constructivism about Reasons." *Oxford Studies in Metaethics* 3: 207–45.
———. 2009. "Evolution and the Normativity of Epistemic Reasons." In supplement, *Canadian Journal of Philosophy* 35: 213–48.
———. 2011. "Mind-Independence without the Mystery: Why Quasi-Realists Can't Have It Both Ways." *Oxford Studies in Metaethics* 6: 1–32.
Talbert, Matthew. 2008. "Blame and Responsiveness to Moral Reasons: Are Psychopaths Blameworthy?" *Pacific Philosophical Quarterly* 89: 516–35.
———. 2012. "Moral Competence, Moral Blame, and Protest." *Journal of Ethics* 16: 89–101.
Todd, Patrick. 2016. "Strawson, Moral Responsibility, and the 'Order of Explanation': An Intervention." *Ethics* 127: 208–40.
Van Hooff, Jan A. R. A. M. 1972. "A Comparative Approach to the Phylogeny of Laughter and Smiling." In *Non-verbal Communication*, ed. R. A. Hinde, 209–41. London: Cambridge University Press.
Van Hooff, Jan A. R. A. M., and Signe Preuschoft. 2003. "Laughter and Smiling: The Intertwining of Nature and Culture." In *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*, ed. Frans B. M. de Waal and Peter L. Tyack, 260–87. Cambridge: Harvard University Press.
Vargas, Manuel. 2013. *Building Better Beings*. Oxford: Oxford University Press.
Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
———. 2014. "Emotions and Relationships: On a Theme from Strawson." *Oxford Studies in Agency and Responsibility* 2: 119–42.
Watson, Gary. 2004. *Agency and Answerability*. Oxford: Oxford University Press.

———. 2014. "Peter Strawson on Responsibility and Sociality." *Oxford Studies in Agency and Responsibility* 2: 15–32.
Wiggins, David. 1987. "A Sensible Subjectivism?" In *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Blackwell.
Williams, Bernard. 1993. *Shame and Necessity*. Berkeley: University of California Press.
Wolf, Susan. 1987. "Sanity and the Metaphysics of Moral Responsibility." In *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman, 46–62. Cambridge: Cambridge University Press.
———. 2001. "The Moral of Moral Luck." *Philosophic Exchange* 31: 4–19.
———. 2011. "Blame, Italian Style." In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, ed. R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 332–47. Oxford: Oxford University Press.
Zimmerman, Michael. 1987. "Luck and Moral Responsibility." *Ethics* 97: 374–86.
———. 1997. "Moral Responsibility and Ignorance." *Ethics* 107: 410–26.
———. 2002. "Taking Luck Seriously." *Journal of Philosophy* 99: 553–76.
———. 2006. "Moral Luck: A Partial Map." *Canadian Journal of Philosophy* 36: 585–608.
———. 2008. *Living with Uncertainty*. Cambridge: Cambridge University Press.
———. 2015. "Moral Luck Reexamined." *Oxford Studies in Agency and Responsibility* 3: 136–59.