

# Guilt without Perceived Wrongdoing\*

Michael Zhao

(forthcoming in *Philosophy and Public Affairs*; please cite published version)

At the age of eighteen, Darin Strauss hit and killed a classmate whose bicycle had swerved, across two lanes of traffic, in front of his car. The police determined that he was not at fault for the accident: after all, he had been driving at a safe speed, and likely couldn't have stopped in time to avoid hitting her. But the accident wracked him with guilt for decades. It wasn't that Strauss blamed himself for his classmate's death. Rather, his guilt was directed toward a "plain, plump truth": "because I'd driven a certain road, someone who had been alive was dead. I had killed someone."<sup>1</sup>

Strauss's case isn't exceptional. We often feel guilt about outcomes without taking ourselves to have been at fault for them. We can feel guilt about harm that befalls a friend as the result of bad advice we gave her, even if we had every reason beforehand to think that the advice was good; about an illness that we passed to a family member, even if we took the necessary precautions against passing it; or about the sacrifices that our parents made on our behalf when we were young children, even if we did not ask those sacrifices of them. In all of these cases, we feel guilt about the *causal* role that we had in a harm to others or in their suffering, even if we were in no way at fault for those things.

---

\*I am grateful to Samuel Scheffler, Sharon Street, Jake Wojtowicz, and two associate editors of *Philosophy and Public Affairs* for their helpful comments on drafts of this paper. I have also presented versions of this paper as talks at NYU, Penn, CUNY Graduate Center, Nebraska, William and Mary, Notre Dame, and NUS, and I am grateful to the audiences at these talks for valuable comments and discussion.

<sup>1</sup>The accident and its toll on his life are the subject of his memoir, *Half a Life* (McSweeney's, 2010). I thank Sharon Street for bringing this example to my attention.

Those readers familiar with the moral luck literature might notice that this case is almost identical to that of the truck driver who hits a child who has run onto the road. Bernard Williams famously described the reaction that we would expect the truck driver to have as "agent-regret" ("Moral Luck," in *Moral Luck* (Cambridge University Press, 1981)), which one might take to be distinct from guilt. I argue in §4 that agent-regret is simply a species of guilt.

Despite the prevalence of this kind of guilt, however, the received account of guilt in the philosophical literature cannot easily make sense of it. According to the received account, the emotion of guilt contains the thought that *one has done something morally wrong*: that, for example, one was morally responsible for some harm to others. If the received account is correct, then one simply cannot feel guilt about outcomes that one does not take oneself to have been at fault for. The account has to explain cases of guilt like Strauss's either as cases *not* of guilt, but of some nearby emotion, or as involving a *mistaken* thought that one was to blame for the bad outcome.

In this paper, I will argue against the received account of guilt, focusing on the case of guilt about being causally responsible—and not morally responsible—for a bad state-of-affairs. I'll start, in §1, by rehearsing the received account. In §2, I'll look at cases of guilt about mere causal responsibility. In §§3–5, I'll argue against responses on behalf of the received account that it can handle these cases. Finally, in §§6–8, I'll defend a novel account of guilt that explains how we can feel guilt about mere causal responsibility for what is bad. On this account, guilt involves the sense that part of the self has been implicated in the occurrence of what is bad.

## 1 The received account

Since guilt is an emotion, I'll begin by discussing the emotions. Throughout, I'll be making the fairly uncontroversial assumption that emotions have some cognitive or representational content: they represent some target object as having a certain property.<sup>2</sup> The emotion of fear, for example, represents some nearby object as posing a danger to the subject: when I am afraid, I think that something nearby can harm me. An instance of an emotion is *fitting* just in case, in that instance, the target object has the property that the emotion represents it as having.<sup>3</sup> So fear when confronted by a rattlesnake in the wild is fitting, since the rattlesnake could kill you; fear when confronted by a toy snake, or a snake on television, is unfitting, since those things are harmless.

This cognitive component of an emotion need not take the form of anything like a consciously-endorsed *belief*, something that the subject might avow ver-

---

<sup>2</sup>See, for example, Patricia Greenspan, *Emotions and Reasons: An Inquiry into Emotional Justification* (Routledge, 1988); Robert Solomon, "On Emotions as Judgments," *American Philosophical Quarterly* 25 (1988): 183–191; Martha Nussbaum, *Upheavals of Thought: The Intelligence of Emotions* (Cambridge University Press, 2001).

<sup>3</sup>See, for example, Justin D'Arms and Daniel Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotion," *Philosophy and Phenomenological Research* 61 (2000): 65–90.

bally. Rather, the thought can simply be what Robert Roberts calls a *construal*, what Tamar Gendler calls an *alief*, or what psychologists like Daniel Kahneman or Jonathan Haidt call an *intuition*, a quasi-perceptual seeming generated through an automatic, unconscious process.<sup>4</sup> So in order to feel fear, you don't need reflectively to believe that some object in front of you poses a danger; all you need is to construe it as dangerous, or alieve or have the intuition that it is dangerous. This is why people often feel fear about things that they reflectively know to be harmless or safe. My mother is afraid whenever she sees a snake on television: she knows that she's not in any danger, but she alieves or has the construal or intuition that she is.

According to the received account of guilt, guilt represents the target object, the subject herself, as having done something morally wrong. This is supposed to be a necessary—in fact, constitutive—condition on feeling guilt: without having this thought, one simply cannot feel guilt. (Different versions of the received account sometimes frame this condition in terms of moral responsibility or blameworthiness, or in terms of some particular wrong. These differences in terminology aren't important for our purposes.)

Again, note that according to the received account, this thought need not take the form of a consciously-endorsed belief. The agent does not need reflectively to believe that he has done anything morally wrong; all he needs is to have the intuition that he has. So think of the case of someone who grew up in an Orthodox Jewish household, where he internalized certain dietary prohibitions: against eating shellfish or pork, or mixing meat with dairy. Later on in life, he comes to reject his upbringing, and no longer reflectively believes that there's anything wrong with these things. But he still might feel a lingering sense of guilt whenever he orders a Cuban sandwich. The received account would explain this by saying that even though he no longer believes that eating pork or mixing meat with dairy is wrong, he still construes them as wrong.<sup>5</sup>

Note that what I'm calling the received account is a descriptive thesis about

---

<sup>4</sup>Robert Roberts, "What an Emotion is: A Sketch," *The Philosophical Review* 97, no. 2 (1988): 183–209; Tamar Gendler, "Alief and Belief," *The Journal of Philosophy* 105, no. 5 (2008): 634–663; Daniel Kahneman, *Thinking Fast and Slow* (Farrar, Straus and Giroux, 2011); Jonathan Haidt, "The emotional dog and its rational tail: A social intuitionist approach to moral judgment," *Psychological Review* 108 (2001): 814–834.

<sup>5</sup>Cf. Rawls's explanation of "residue guilt" in ch. VIII, §73 of his *Theory of Justice*, Revised Edition (Belknap, 1999). Throughout, when I say that a subject *has the thought* that some object has or *takes* it to have some property, I mean that he *construes* that object as having that property: having a quasi-perceptual seeming, rather than a fully-fledged belief, that the object has the property.

the nature of guilt. It doesn't have any implications for when one *deserves* or when it's *expedient* to experience guilt. It does imply a claim about when it's *fitting* to feel guilt, namely, when the subject really has done something wrong. But claims about fittingness don't imply claims about other kinds of normativity. Again, the claim that an emotion is fitting is simply the claim that the cognitive component of the emotion is correct: that its target object has the property that the emotion represents it as having. And this doesn't automatically settle the question of whether it would be prudent or morally appropriate to have the emotion.

Now, the received account does tend to encourage conflation of the fittingness of guilt with its moral deservedness. After all, it says that guilt is fitting so long as the agent has done something morally wrong. And because it is a painful emotion, it is natural to think that it is morally deserved in exactly those cases: the pain of guilt is deserved as a kind of self-punishment when one has done something wrong. But we should bear in mind that the question of when guilt is fitting, in the sense of its cognitive component being correct, is conceptually separate from the question of when it is morally deserved, perhaps as a kind of self-punishment for what one has done.

My account of guilt will challenge the cognitive component that the received account assigns to guilt; I'll argue that guilt represents the subject as having a different property. My account, then, will also be descriptive: in claiming that guilt takes a different thought, I'm claiming only that guilt is *possible* whenever the subject has that thought, without making any claim about when guilt is expedient or morally deserved. To the extent that these descriptive claims about the cognitive component of guilt imply claims about its fittingness, my account of guilt will allow for the possibility that guilt is fitting without being morally deserved (although, as I'll show, it is not committed to this possibility). But there is nothing objectionable about this: I can fittingly feel guilt without deserving to be wracked by its pain, just as I can fittingly feel grief over the loss of a family member without deserving to be wracked by *its* pain.

The received account of guilt may seem obviously correct; after all, it might just seem platitudinous to say that guilt is a response to one's own wrongdoing. To make things seem almost tautological, *guilt* is the emotion that we have when we think that we are *guilty* of wrongdoing. And many philosophers have casually assumed something like the received account in their discussions of guilt, treating it as self-evident rather than contentious enough to require justification. John Rawls writes, for example, that one who feels guilty does so "because he

has acted contrary to his sense of right and justice.”<sup>6</sup> R. E. Lamb writes, “we can (intelligibly be said) to *feel* guilty about *x* only if we *believe* that we are, in a moral sense, *responsible* for *x*.”<sup>7</sup> Walter Sinnott-Armstrong writes, “Guilt occurs when one’s bad feelings are caused by one’s belief that one’s act was morally wrong.”<sup>8</sup> Martha Nussbaum writes, “Guilt is a type of self-punishing anger, reacting to the perception that one has done a wrong or a harm.”<sup>9</sup> Daniel Jacobson writes that “guilt is typically taken to be rationally restricted to the voluntary—or, at any rate, the blameworthy.”<sup>10</sup> (He’s using “rational” to mean fitting here, so it follows from what he says that guilt is possible only if one takes oneself to be blameworthy for something.) And finally, David Sussman writes that guilt “involve[s] a pained awareness of one’s own wrongdoing.”<sup>11</sup>

I want to give special attention to remarks that Bernard Williams made about guilt at various points, since he says things that point us toward what I take to be the correct account of guilt. Williams endorsed (somewhat inconsistently, as we’ll see) the received account of guilt in his discussions of moral responsibility. Toward the end of *Ethics and the Limits of Philosophy*, when he introduces the idea of the “morality system,” one that structures morality around the notion of moral responsibility, he writes that “remorse or self-reproach or guilt ... is the characteristic first-personal reaction within the system.”<sup>12</sup> This seems to imply that guilt requires the thought that one was morally responsible for some bad outcome. Later, in *Shame and Necessity*, he writes that “what arouses guilt in an agent is an act or omission of a sort that typically elicits from other people anger, resentment, or indignation,” and these are attitudes that one has in reaction to intentional harm, injustice, or other wrongdoing.<sup>13</sup>

One of Williams’s aims was to build up guilt as a foil to the emotion of shame, which doesn’t require any notion of moral responsibility, and which Williams took to be more central in our moral lives than we thought. This was part of his larger project of attacking the morality system, showing that our moral lives are

---

<sup>6</sup>*A Theory of Justice*, 391.

<sup>7</sup>“Guilt, Shame, and Morality,” *Philosophy and Phenomenological Review* 43, no. 3 (1983): 340. Italics his.

<sup>8</sup>“You Ought to Feel Ashamed of Yourself (when You Violate an Imperfect Moral Obligation),” *Philosophical Issues* 15 (2005): 200.

<sup>9</sup>*Hiding from Humanity: Disgust, Shame, and the Law* (Princeton University Press, 2006), 207.

<sup>10</sup>“Regret, Agency, and Error,” *Oxford Studies in Agency and Responsibility* 1 (2013): 107.

<sup>11</sup>“Is Agent-Regret Rational?” *Ethics* 128, no. 4 (2018): 792.

<sup>12</sup>*Ethics and the Limits of Philosophy* (Harvard University Press, 1985), 177.

<sup>13</sup>*Shame and Necessity* (University of California Press, 1992), 89.

less dependent on a Kantian notion of moral responsibility, hence allow for more luck, than we might think. But if guilt does not require moral responsibility, as I'll argue, then it was a hostage that Williams need not have taken: the conditions for the fittingness of guilt are susceptible to a good deal of luck as well.

## 2 Guilt about causal responsibility

Even though the received account of guilt may seem obviously right at first glance, when we look at the wide range of cases in which people experience what they describe as "guilt," the correctness of the account becomes much less obvious. After all, people feel guilt about wrongdoing that a person or group with whom they identify performed, even if they were individually innocent: think about the guilt that many Germans born after WWII feel about the Holocaust. People feel guilt about violating certain non-moral norms, like those about dieting or exercising or not being lazy. And they feel guilt about undeservedly doing better than others, like surviving an accident or atrocity that those around them did not. In none of these cases do the subjects necessarily take themselves to have done anything morally wrong.<sup>14</sup>

In this section, I'll focus on a fourth kind of guilt without perceived wrongdoing: guilt about mere *causal* responsibility for what is bad. The existence of guilt about outcomes for which we hold ourselves merely causally, and not morally, responsible for is a problem for the received account. In the next few sections, I'll consider and argue against some responses on behalf of the received account.

Let me start, though, with some examples of mere causal responsibility for a bad state-of-affairs that might cause guilt. First, take an example that is relevantly identical to the case of Darin Strauss, but likely more familiar to philosophical readers. This is the example of the blameless truck driver, made famous by Bernard Williams's and Thomas Nagel's discussions of moral luck:

*(Truck driver.)* A truck driver, through no fault of his own, hits and kills a child who has run onto the road.

Next, consider an example from the life of Elvis Presley:

---

<sup>14</sup>At this point, one might wonder to what extent the dispute is verbal. Perhaps proponents of the received account are simply using "guilt" in a way that it does not apply to these cases, while ordinary speakers are using it in a way that does. But, first, verbal disputes are still disputes; if proponents of the received account are using "guilt" in a way that does not line up with ordinary usage, they should make it explicit that their use is stipulative. And second, even if the dispute is verbal, the task of providing an account of the emotion that ordinary people describe as "guilt" still remains open, which is the task that I am undertaking.

(*Elvis.*) Elvis was born with a stillborn twin. As an adult, he confided to a friend that he thought that he had caused his twin to die while in the womb, by absorbing more than his share of nutrients.

And finally, consider the following, also taken from a real event:

(*Rescuer.*) A man was driving to work when his car unexpectedly hit an icy patch on a bridge and flipped over. Local firefighters were sent to pull him out from the wreckage. In the process of reaching him, one of the firefighters slipped and fell through a gap in the roadway, plummeting to his death.<sup>15</sup>

In all of these cases, guilt is a natural (although perhaps not universal) reaction. Drivers who cause injury or death to others often feel guilt, even if they know that they were not at fault. Elvis apparently felt guilt throughout his life for the role he took himself to have had in his twin's death. And those whose rescue leads to the death of a rescuer frequently feel guilt too; the subject of the last example said that he "immediately felt guilty" when he learned of the firefighter's death.

In all of these cases, the subjects take themselves to be causally responsible for some bad state-of-affairs. After all, they take a counterfactual condition between them and that state-of-affairs to hold: for some salient act or omission or event, had they not performed that act or omitted that omission or experienced that event, the bad state-of-affairs would not have obtained.<sup>16</sup> If the truck driver hadn't been driving on that road, the child would still be alive. If Elvis hadn't absorbed more than his share of nutrients, (by his own reasoning) his twin would still be alive. (There's a separate question of whether the scientific basis for Elvis's belief is sound; what's relevant here is just that Elvis *took* himself to be causally responsible for his twin's death.) If the man hadn't been in the wrecked car, the firefighter would still be alive. In each of these cases, the subject thinks that the bad outcome obtained in part *because of* him; he might think of himself as part of the *reason* that the outcome obtained.

---

<sup>15</sup>See "Man whom firefighter died trying to save feels 'guilty,'" *New York Post*, January 7, 2019.

<sup>16</sup>This counterfactual ("but-for") condition is not *necessary* for causal responsibility: after all, even if it were true that if Oswald hadn't shot Kennedy, someone else would have, Oswald would still be causally responsible for Kennedy's death. But the condition is *sufficient* for causal responsibility.

But it also seems unlikely, at least *prima facie*, that the subjects take themselves to be *morally* responsible for the tragic outcome in any of these cases. After all, there are two analytic conditions on moral responsibility for an outcome. First, there is a *voluntariness* condition: the subject must have been acting freely; and second, there is an *epistemic* condition: the subject must have been able reasonably to foresee that the outcome would (or could, with some non-negligible probability) occur as the result of her actions. In each of the examples, it seems unlikely that the subject would take both conditions to hold. In *truck driver* and *rescuer*, it seems implausible that the subjects take the epistemic condition to hold: the truck driver knows he could not reasonably have foreseen that his driving on that road would cause the death of a child, and the man driving to work knows he could not have reasonably foreseen that his driving on the bridge would lead to the death of a firefighter. And in *Elvis*, it seems unlikely that Elvis took the voluntariness condition to be satisfied: after all, Elvis knows he wasn't doing *anything* voluntarily that brought about the death of his twin. He wasn't acting at all while in the womb; he was simply undergoing the metabolic processes that a fetus undergoes.<sup>17</sup> It seems unlikely that the subjects would say that they did anything *wrong*.<sup>18</sup> These are cases, then, of guilt about mere *causal* (and not moral) responsibility for some regrettable state-of-affairs.

---

<sup>17</sup>I address the response that, even if the subjects *know* that they weren't morally responsible, they might nonetheless *construe* themselves as morally responsible in §5.

<sup>18</sup>The reader might demur at this last claim. After all, one might think that killing an innocent person is wrong, and the truck driver (no matter how unintentionally) has violated a moral obligation in doing so. Paulina Sliwa, for example, makes this objection: "Still, it is false that [the truck driver] has done nothing wrong: he killed a child. The killing was neither intentional, reckless, nor negligent. But our moral obligation is not to kill, not merely to refrain from killing intentionally, recklessly, or negligently" ("The Power of Excuses," *Philosophy and Public Affairs* 47, no. 1: 67). Perhaps we could distinguish between wrongdoing as a mere *actus reus*, violating a moral obligation in fact, from wrongdoing that includes a *mens rea*, violating a moral obligation intentionally, knowingly, or through negligence. If we defined the received account of guilt in terms of wrongdoing as an *actus reus*, then the truck driver example would not be a counterexample to it.

In response to this suggestion, however, note first that it would be in tension with the alleged function of guilt as a form of self-punishment. After all, punishment is fitting only when one has harmed another with the relevant *mens rea*: it would be unjust to punish the truck driver, when he neither intended nor could have foreseen the harm to the child. If guilt is the internalized version of the punishment that others might dole out on the subject, it would similarly be unfitting in cases of wrongdoing as a mere *actus reus*. And second, even if this response works for the truck driver example, it cannot work for the Elvis and rescuer examples. After all, in those cases, the subjects weren't *doing* anything—they weren't acting—so, *a fortiori*, they couldn't have been doing anything wrong in either sense of "wrongdoing."



Before moving on, I want to mention a few important points. First, perhaps not everyone would feel guilt if placed in these cases. There are many who feel guilt only when they take themselves to have been at fault for the bad outcome. I'll treat it as a datum that mere causal responsibility for what is bad is enough to elicit guilt for some, but not for others. Given that the received account takes the thought that one has done something wrong to be a component of guilt, the mere fact that many feel guilt in cases in which they do not take themselves to have done wrong still makes trouble for the received account. The account of guilt that I defend will be compatible with the diversity in the eliciting cases of guilt for different people, and will say in virtue of what some feel guilt about being merely causally responsible while others do not.

Second, I want to note that causal responsibility is a broad genus, and there are three species of causal responsibility in play in the examples above. First, in the truck driver example, the subject is causally responsible for the death of the child because it is the consequence of one of his *actions*, driving on that road. Another way of putting this is that killing the child was one of the driver's actions, although it was unintentional under that description. Since this form of responsibility involves the subject's agency, I'll call it *agential* responsibility. Next, in *Elvis*, the subject is the proximate cause of some harm, but in a way that does not involve his agency; he is simply the instrument of the harm, or the vector through which the harm is transmitted. I'll call this form of causal responsibility *instrumental*.<sup>19</sup> And finally, in *rescuer*, the subject is causally responsible for the outcome because *someone else* brings it about, either in order to have an effect on the subject, or as a byproduct of something else done to have an effect on the subject. I'll call this form of responsibility *indirect*.<sup>20</sup>

Third, although I'll talk about guilt as a response to one's involvement in bad (or unfortunate, or regrettable) outcomes, I want to note that there are different ways in which a state-of-affairs can be bad, involvement in only one kind of which can be an occasion for guilt. I'll say that a state-of-affairs is *morally bad* if there is moral reason not to bring that state-of-affairs about; in contrast, it is non-morally bad if there is some other kind of reason (like a prudential one) not to bring it about. As one example of this distinction, most of us regard harms

---

<sup>19</sup>As some other forms of guilt about instrumental responsibility, consider the guilt of someone who catches a deadly disease and survives, but passes it on to a family member, who dies from it; or the guilt that a parent might feel about giving birth to a child with physical deformities.

<sup>20</sup>As some other examples: your parents sacrifice their lives in order to save yours; an assassin accidentally shoots your daughter while attempting to kill you; your husband organizes an assault on a rival skater to improve your chances of success.

to others as morally bad, whereas we regard most harms to ourselves as non-morally bad: there is *moral* reason not to risk serious injury to others, but only *prudential* reason not to risk serious injury to ourselves. When I write that guilt is a response to involvement in bad states-of-affairs, I'm using "bad" in the sense of morally bad; in other words, we experience guilt over the outcomes that we take there to have been moral reason not to bring about, like ones that feature harm to or the deaths of others. After all, we don't experience guilt about what we regard as non-morally bad: the decision not to pursue one's dreams might be, decades down the line, an occasion for deep regret, but not for guilt.

Finally, one might wonder if any degree of causal responsibility for anything bad is sufficient to elicit guilt. Paulina Sliwa notes that, in the truck driver case, many of the agents who had some degree of causal responsibility for the child's death, like the truck manufacturer or the manager who sent the truck driver on the route, do not seem like possible subjects of sorrow or guilt.<sup>21</sup> More generally, we are all causally responsible, through long and torturous causal chains, for a whole host of outcomes simply through the seemingly innocent things we do on a daily basis. Some of these outcomes are terrible, but we do not feel guilt about them.

I acknowledge that not every instance of causal responsibility for a bad will elicit guilt. First, we are often ignorant of our causal responsibility for bad outcomes, so cannot feel guilt about our role in bringing those outcomes about. It is unlikely that the truck manufacturer, for example, knows about its causal responsibility for the death of the child. Again, what matters for the possibility of guilt is not that one *is* causally responsible for the bad outcome, but that one *takes oneself to be*. Second, our degree of causal involvement in an outcome may be limited enough to attenuate any feelings of guilt: to the extent that an outcome required an entire network of variables, in which our actions are only one small piece, to be set in a particular way, our causal responsibility for the outcome may be diffuse enough not to generate any strong self-directed emotions. The manager at the trucking company, for example, might take his involvement in the child's death to be too weak and indirect to feel guilt about it. In order to feel guilt about something that one takes oneself to be causally responsible for, one's perceived degree of causal involvement in that outcome must be taken to be beyond some threshold of significance.

---

<sup>21</sup>"The Power of Excuses," 67.

### 3 Are these cases of unfitting guilt?

These examples of guilt without perceived wrongdoing pose a problem for the received account, according to which guilt is possible only if the subject takes himself to have done something wrong. In order to defend it, its proponent has two options: first, he can deny that these are genuine cases of guilt; second, he can find a way to attribute to the subjects the mistaken thought that they did something wrong. I'll consider and argue against those responses in §§4–5. In this section, I want to discuss a preliminary response on behalf of the received account, one that does not purport to diagnose where my argument against the received account has gone wrong, but that nonetheless provides some evidence in support of that account.

One natural reaction to the cases that I've discussed might be that they are ones in which guilt does not seem *fitting*. After all, we do tell people like the subjects of the examples that we discussed that they *shouldn't* feel guilt, since the outcome wasn't their fault. The subject of *rescuer*, whose rescue led to the death of a firefighter, recalls that other firefighters tried to comfort him after the death of their comrade by telling him that the death wasn't his fault. Given that the rejection of the received account implies that guilt can be fitting even when the subject hasn't done anything wrong, our reaction that these cases of guilt are unfitting supports the received account.

But it is unclear how we are using “should” when we tell people that they shouldn't feel guilt: in the sense of fittingness, or in the sense of moral appropriateness. First, as we mentioned, an emotion can be fitting without there being moral reason for the subject to have the emotion: it is fitting for someone who lost a family member to feel grief, although no one would suggest that there is moral reason for her to feel guilt—that, for example, she deserves to feel grief. If what we mean when we tell people that they shouldn't feel guilt when they weren't at fault for the outcome is just that there is no *moral* reason for them to feel guilt, that is consistent with thinking that guilt is nonetheless *fitting* in those cases.

Second, even if we are using “should” in the sense of fittingness, it is unclear how much we can glean about the fittingness of guilt from the mere practice of telling people that they shouldn't feel guilt when they weren't at fault. For one, people often react to such attempts to comfort them in a way that reveals that they take their own guilt to be fitting. As one (admittedly anecdotal) piece of evidence for this claim, consider the case of a woman interviewed in a *New Yorker*

piece on drivers who accidentally caused the deaths of others.<sup>22</sup> This woman, momentarily blinded by a gleam of light while driving, hit and killed a motorcyclist; understandably, she felt guilt for years after the accident. She reports that friends kept trying to comfort her by telling her things like “it was just an accident” or “it wasn’t your fault,” and writes of her impatience at hearing these attempts to console her: “Yes, it was an accident, . . . but, at the end of the day, I hit him, I took his life . . . Why does nobody understand this?” So at least in this case, the subject seems to think *not* that her guilt is unfitting, because it *mistakenly* represents her as having been at fault for the motorcyclist’s death, but that her guilt *is* fitting, because it *correctly* represents her as having killed the motorcyclist. So the evidence from the practice of telling people who are merely causally responsible for harms to others that they shouldn’t feel guilt doesn’t seem unequivocally to support the idea that their guilt is unfitting.

As an additional piece of evidence that guilt can be fitting even when the subject hasn’t done anything wrong, we would find there to be something worrying or chilling about someone who didn’t feel guilt in cases of causal responsibility: we’d think that there’s something emotionally defective about a truck driver who feels no guilt over the death of the child. (Williams, in discussing the truck driver case, writes that “some doubt would be felt about a driver who too blandly or readily” rid himself of all but the emotions that a spectator to the accident would have.<sup>23</sup>) One natural explanation for what is defective about such a subject is that he is making a quasi-perceptual error, failing to see that he has whatever property guilt represents him as having. If this explanation is correct, so the driver really has the property that guilt represents him as having, then guilt is a fitting response to his situation.

Finally, one might concede that these cases of guilt *are* unfitting, but nonetheless want to preserve their possibility without requiring that the subjects mistakenly construe themselves as having done wrong. The received account of guilt cannot do this, but I will show (at the end of §8) how my account can.

#### 4 Are these even cases of guilt?

I’ve offered some reason to reject a preliminary response on behalf of the received account, that guilt seems unfitting when the subject hasn’t done anything wrong. In this section, I will discuss another response, which purports to identify where

---

<sup>22</sup>Alice Gregory, “The Sorrow and Shame of the Accidental Killer,” *The New Yorker*, September 11, 2017.

<sup>23</sup>“Moral Luck,” 28.

my argument against the received account has gone wrong: that what the subjects in the examples are feeling is not *guilt*, but rather a separate emotion.

One might think that what I've called "guilt" in the case of mere causal responsibility isn't guilt at all. Readers familiar with the moral luck literature might be aware that Williams actually uses the term "agent-regret" rather than "guilt" to describe the truck driver's reaction.<sup>24</sup> And one might think that what people are feeling in these cases is really a separate emotion, agent-regret, which doesn't require perceived culpability in the way that guilt does. After all, agent-regret is supposed to have as its "constitutive thought" that I did something that turned out to have bad consequences. If what the people in the examples are experiencing is agent-regret rather than guilt, then these cases don't pose a problem for the received account of guilt.<sup>25</sup>

But this response is inadequate. First, it doesn't apply to all cases of guilt about causal responsibility. After all, the emotion appealed to is called "*agent-regret*," and it's regret about one's agency being implicated in badness. But we noted that there are various kinds of causal responsibility in play in the examples above, two of which don't involve agency at all. In *Elvis* and *rescuer*, the subjects needn't have *done* anything voluntarily that brought about the bad outcome. Rather, they were causally responsible for it in some other way, either instrumentally or indirectly.

Second, the response of distinguishing agent-regret from guilt seems *ad hoc*. Consider that guilt and agent-regret are supposed to be identical in terms of their phenomenology and motivational pattern: they both involve a feeling of pain or anxiety, and they both generate a motivation to compensate those harmed. Williams himself notes that a desire to recompense the victim is a typical component of agent-regret: the truck driver, for example, might offer something to the parents of the dead child as a gesture of restitution. In these respects, guilt and agent-regret are identical, which suggests that any distinction between them, in terms of their cognitive components, is likely to be artificial. (From here on,

---

<sup>24</sup>"Moral Luck."

<sup>25</sup>In the discussion of the truck driver example, it is often assumed that what the driver is feeling cannot be guilt if he takes himself to be innocent. David Enoch writes, for example, "if the driver believes herself to be somehow *guilty* of the death (perhaps she was driving terribly negligently), then guilt may be the thing she can and the bystander typically cannot feel"—the implicature being that guilt cannot be what the driver feels if she does not take herself to be guilty for the child's death ("Being Responsible, Taking Responsibility, and Penumbra Agency," in Ulrike Heuer and Gerald Lang, eds., *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams* (Oxford University Press, 2012), 96).

I'll treat agent-regret as a species of guilt.)

Third, people in these kinds of cases often describe themselves as feeling guilt. Strauss, reflecting on the toll that the accident took on his life, writes, "That's the force of guilt for you." In the same vein, there's a website, "Accidental Impacts,"<sup>26</sup> for people to share stories of automobile accidents that they caused. Some of the postings include lines like, "I feel guilty and wish it would have been me [who had died]," "I felt and feel so guilty about the death of their child," and, "It has totally ruined my life, I am highly stressed, irrational and irritable with a heavy heart and a terrible sense of loss and guilt." The motorists in these cases were found not to have been at fault for the accidents, but they still feel what they describe as guilt for their role in them.

Now, Williams might have had strategic reasons for coining a new term to describe the truck driver's reaction instead of using "guilt": since he has built up guilt as a reaction distinctive of the morality system, he does not want to detach it from moral responsibility. This forces him to introduce an emotion that is phenomenologically and motivationally identical to guilt, but that does not require an attribution of moral responsibility to oneself. But it turns out that Williams equivocates on whether guilt requires this self-attribution of moral responsibility; in *Shame and Necessity*, he admits that "agent-regret ... can be psychologically and structurally a manifestation of guilt."<sup>27</sup>

What if the emotion that the subjects have is *shame* instead? After all, shame and guilt are among the most popular negative self-reactive attitudes. And unlike guilt (according to the received account), shame does not require our being morally responsible for what elicits it: we can feel shame about our character, our background, or our appearance, even if we were not responsible for any of these things.

Unfortunately, it does not seem plausible that what the subjects are experiencing in these cases is shame either. Philosophers typically take shame to involve the experience of oneself as having certain features, like character traits, that one considers undesirable. Walter Sinnott-Armstrong, for example, contrasts shame with guilt by saying that, while guilt requires a perception of oneself as having acted wrongly, "in shame, the bad feeling is caused by one's belief

---

<sup>26</sup><https://accidentalimpacts.org>.

<sup>27</sup>*Shame and Necessity*, 93. Another reading of Williams, however, is that he has two notions of guilt in mind: a restrictive one, on which it is a response to wrongdoing in particular, and an expansive one, on which agent-regret counts as a form of guilt. Thanks to Jake Wojtowicz for offering this suggestion to me.

that one's self or one's character is defective in a more general way."<sup>28</sup> In a case in which the bad outcome *was* the product of the subject's negligence, the subject might feel shame about his own negligence, a defective character trait made salient by its role in the terrible outcome. In contrast, in the case of *truck driver* and *rescuer*, neither subject was acting negligently, so the experience of themselves as negligent could not be the occasion for whatever self-reactive attitude they have. More generally, because the outcome cannot be attributed to anything like a stable character trait of the subject's, it is unclear which defective trait the subjects experience themselves as having. So there is not a strong case that the subjects are feeling shame instead of guilt.

## 5 Do the subjects really take themselves to be blameless?

So much for the response that what the subjects are feeling in these examples is not guilt, but rather a separate emotion. A final possible response on behalf of the received account is to explain these cases by finding some way to attribute to the subjects the *mistaken* thought that they in fact were to blame for the bad state-of-affairs, hence that they really did something morally wrong.

There are a number of ways in which this response could go. First, we might think that the subjects cannot be *certain* that they were not at fault, and feel guilt on the basis of having some degree of belief that they were. Consider, for example, *truck driver* again. Even if the truck driver was completely blameless, we might think that he can't be certain of that fact. After all, lots of accidents are ambiguous, and maybe there's no way for him to know that he wasn't being too negligent while driving. Given that the situation was ambiguous, he might choose to fixate on the worst interpretation of the event, that he was being too negligent. He might think, in self-reproach, "I should've been going a little bit slower! I should've been paying a little more attention!" And even if the truck driver lacks an *outright* belief that he was responsible, he might nonetheless have some non-negligible credence that he was. So we might explain the truck driver's feeling guilt in terms of his mistaken thought that he was, or might have been, at fault for the accident.

We might even think that, when the situation is ambiguous, there's something morally admirable about erring on the side of assuming responsibility for too much rather than for too little. We might think that there's something admirable in the truck driver's feeling of guilt, even if it is unfitting. This line has been made in the literature on moral luck. Daniel Jacobson writes, for example,

---

<sup>28</sup>"Ashamed," 200.

People tend to beat a hasty retreat from responsibility for bad outcomes, often under the cover of the passive voice. In light of this dubious tendency, we should acknowledge that there are grounds for self-scrutiny whenever one is the causal agent of a bad outcome. An admirable person in a real-life situation would be prone to search for grounds of self-reproach: he cannot help himself to the stipulation that he is blameless.<sup>29</sup>

Brian Rosebury makes a similar point, that having some doubt about his own innocence is indicative of a praiseworthy awareness of human epistemic fallibility.<sup>30</sup>

But it's unclear how this response would generalize to some of the other cases of guilt about mere causal responsibility that we discussed. After all, the explanation that I've just given of how the truck driver could mistakenly take himself to have been at fault involved a tendency to focus on the worst interpretation of ambiguous cases, or to assume more responsibility rather than less in them. And it's unclear how an analogous explanation would go in *Elvis*. Unlike *truck driver*, in which there are a number of different interpretations of the event that are possible from the subject's epistemic perspective, Elvis is unambiguous; from Elvis's perspective, there is no interpretation of the events leading to his twin brother's death on which he was morally responsible for that outcome. Given this fact, positing any tendency to assume more moral responsibility than one actually bears in ambiguous situations simply does not apply in the case of Elvis. Rather, what seems to elicit Elvis's guilt is the mere thought that he was the instrument for his twin's death—the thing through which his twin's death occurred—which does not require his culpability.<sup>31</sup>

---

<sup>29</sup>"Regret, Agency, and Error," 114.

<sup>30</sup>"Moral Responsibility and Moral Luck," *Philosophical Review* 104, no. 4 (1995): 499–524.

<sup>31</sup>A reviewer has helpfully pointed out that the Elvis example is drawn from a culture in which the notion of "original sin" is popular, according to which people have quasi-agential faults even before they have full-fledged agency. If Elvis took his absorbing more than his share of nutrients while in the womb as a reflection of his sinful nature, that could help explain his feeling of guilt about his twin's death without requiring the rejection of the received account.

In response, I acknowledge that this might explain other negative self-directed attitudes, such as shame; but it does not seem like a good explanation of guilt. As we mentioned, shame is often taken to be a reaction to the undesirability of one's traits or other characteristics. Experiencing one's nature as greedy or sinful might therefore be an occasion for shame. According to the received account, however, guilt requires not just that the subject possess certain faults associated with agency; the subject must actually have acted freely in a way that he was in a position to know that he should not have acted in. Even if Elvis had taken himself to have various quasi-agential



A second way of making the response that the subjects are mistakenly taking themselves to have been at fault is as follows. Even if the subjects are in a position to *know* that they were not at fault, why not think that they nonetheless *construe* themselves as having been morally responsible for the outcome? After all, we know that emotions, like perceptions, can be *recalcitrant*, persisting even when the subject knows that they are mistaken. Even when I know that the lines in the Müller-Lyer illusion have the same length, I still perceive one as longer than the other; similarly, even though my mother *knows* that the snake on television presents no danger to her, she still construes it as dangerous, and is afraid of it. The mere fact, then, that the subjects know that they are not morally responsible for the outcome does not show that they do not mistakenly construe themselves as morally responsible, and feel guilt as a result of that mistaken construal.

In response to this objection, consider what explains recalcitrant emotions in general. A natural way to do this would be to draw a parallel to the case of perceptual illusions: here, one popular explanation for their existence is in terms of our subconsciously employing some assumption that is usually reliable, but that generates a false perception in a minority of cases.<sup>32</sup> The assumptions that we rely on must be typically reliable; otherwise, there is no explanation for why we would rely on them in the first place.<sup>33</sup> Consider, for example, the Ames room, a room that has been cleverly designed to give the *appearance* of being rectangular in size from where the subject stands, but that is actually trapezoidal, with one corner of the far end much closer to the subject than the other corner. When two objects of identical size (like a pair of identical twins) are placed at the far corners, we construe one as being much larger than the other, even though we know that they are the same size. What explains the illusion that one object is much larger than the other is an assumption that our brain makes that rooms that look to be

---

defects while still a fetus, that does not suggest an obvious explanation for why he mistakenly construed himself as having acted freely while still a fetus.

<sup>32</sup>The idea that perceptual illusions (and perceptions, more generally) are the result of unconscious assumptions and inferences dates back to Hermann von Helmholtz's *Treatise on Physiological Optics*. For a more recent exposition, see ch. 4 of Steven Pinker, *How the Mind Works* (Norton, 1997).

<sup>33</sup>This is a bit simplistic; there is a trade-off between reliability, ease of use, and ability to avoid costly errors in the assumptions that we subconsciously make. The assumption that snake-like objects are dangerous may not be very reliable: after all, a large proportion of things shaped like snakes (like long gourd vegetables or garden hoses, or even many species of snakes) are totally harmless. But because the cost of failing to construe a deadly snake as dangerous is so high, we employ this assumption, flinching whenever we see a snake-like object, even though it is not very reliable.

rectangular are rectangular, so that we construe the objects at the far corners as being equidistant from us, hence different in size. This is a reliable assumption to make, given that the vast majority of such rooms that we encounter are not cleverly disguised rooms of other shapes; but in the case of the Ames room, the assumption fails, and a perceptual illusion results.

We can naturally extend this explanation of perceptual illusions to the case of recalcitrant emotions: these emotions occur when we subconsciously employ some assumption that is usually reliable, but that goes wrong in a minority of cases, generating a mistaken construal.<sup>34</sup> As one example, we seem to subconsciously employ the assumption that violent motion is an indication of danger. This assumption is typically reliable, especially in the environment in which many of our unconscious assumptions evolved: the vast majority of cases of violent motion would have been correlated with things (like earthquakes or landslides) that were dangerous to us. But as a result of subconsciously employing this assumption, we perceive violent motions that we know to be perfectly harmless (like turbulence on an airplane) as dangerous, which creates fear even when we know that it is unfitting.

So in order to make the case convincingly that the subjects in these cases are mistakenly construing themselves as morally responsible, even when they're in a position to know that they aren't, we have to be able to find an assumption that it is plausible that they are subconsciously employing, and that would lead them to construe themselves as morally responsible even when they aren't. It is unclear, however, what this assumption would be. Consider, for example, the assumption that cases in which we are causally responsible for an outcome are also cases in which we are morally responsible for that outcome. If we subconsciously employed such an assumption, it would explain our construing ourselves as morally responsible for an outcome in cases in which we are only causally responsible, even if we are in a position to know better, and this would explain the subjects' feeling guilt in the examples we discussed. The problem with this explanation, however, is that it seems implausible that we would make such an assumption in the first place. After all, I mentioned that the assumptions that we subconsciously employ are typically *reliable*: the proportion of cases in which the assumption is false is low. And in the case of this particular assumption, the proportion of false cases is quite high: the range of outcomes for which we are causally

---

<sup>34</sup>On the analogy between perceptual illusions and recalcitrant emotions, see Justin D'Arms and Daniel Jacobson, "The Significance of Recalcitrant Emotion," *Royal Institute of Philosophy Supplement* 52 (2003): 127–145; although cf. Michael Brady, "Recalcitrant Emotions and Visual Illusions," *American Philosophical Quarterly* 44, no. 3 (2007): 273–284.

responsible far outstrips the range for which we are morally responsible, since we're often not at fault for the harms that we cause. So it is implausible that we would subconsciously employ this highly general assumption, which is false in a large proportion of cases. The burden of proof is on the defender of the received account to identify an assumption that the subject could plausibly be relying on subconsciously that would generate the mistaken construal in these cases, and it is unclear which assumption they could appeal to. It is unlikely, then, that the subjects feel guilt because they mistakenly construe themselves as morally responsible for the bad outcome.<sup>35</sup>

## 6 Why do we feel guilt about agential responsibility?

So far, we've considered and rejected a number of possible responses on behalf of the received account. The failure of these response should make us pessimistic about the received account's ability to explain cases of guilt in which the subject takes himself to have been merely causally responsible for some bad state-of-

---

<sup>35</sup>Finally, a variant on the strategy pursued in this section is to argue that the subjects feel guilt not toward the tragic outcome itself, but toward *their own attitude* toward that outcome, which they may regard as morally wrong to feel. Consider that, in some cases, the survival of the subject was possible only at the cost of someone else's life. In the world of the novel *Sophie's Choice*, for example, Sophie's son may realize that his survival was possible only because of the death of his sister; had she not died, he would have instead.

Now, in these cases, the subject often *affirms* his own survival: he is glad rather than regretful to have survived, and cannot help but want things to go the same way, if they had to happen again. Many think, however, that in affirming an event, one is committed to affirming all of the events that were causally necessary for that event to occur. So in affirming their own survival, the subjects are committed to affirming the tragedy that befell others, since that tragedy was necessary for their survival. (R. Jay Wallace has an extended discussion of this dynamic in his *View from Here: On Affirmation, Attachment, and the Limits of Regret* (Oxford University Press, 2013).) But they may regard it as morally wrong to affirm that terrible event, so also regard it as wrong to affirm their own survival, and feel guilt about the attitude of affirmation that they do feel about having survived.

Whatever can be said about *these* kinds of cases, however, the cases that I described above do not have the same kind of causal structure. In those cases, the tragic outcome was not necessary for the survival of the subject: the truck driver knows that both he and the child could have survived, if only he had stopped in time; Elvis believes that both he and his twin brother could have survived, if only he had absorbed fewer nutrients while in the womb; the man in the wrecked car knows that both he and the firefighter could have survived, if only he had missed the icy patch. If we are committed to affirming only those events that were *necessary* for whatever events we affirm, and not all of the events that preceded the events that we affirm, then these subjects need not affirm the tragic event as a requirement on affirming their own survival, and it becomes implausible to explain their guilt as guilt about their affirming what is bad.

affairs.

In the remainder of this paper, I will introduce and defend a novel account of guilt. In light of the preceding discussion, an account of guilt should satisfy several desiderata. First, it must explain the familiar cases of guilt: those in which we feel guilt about what we take ourselves to be morally responsible for. Second, it must explain why many feel guilt about what they take themselves to be merely causally responsible for. Third, the account has to be open-ended enough to allow for other forms of guilt without wrongdoing that aren't our focus, like vicarious or survivor guilt. Fourth, the account must not be too permissive; in particular, it should not allow for the possibility that mere spectators to some tragic event feel guilt about it (to the extent that they do not think that they could have prevented the event).

These four desiderata are conditions on the extensional adequacy of an account of guilt. Finally, we need an additional constraint: our account of guilt must make the *self-directed* nature of the emotion clear. It must show, for example, how guilt differs in its character from emotions available to a mere spectator, like dismay or garden-variety regret. Williams writes that what distinguishes the agent-regret that the driver feels and the emotions available to a spectator is that the sense that the driver “stands in a special relation to the happening,”<sup>36</sup> a relation that does not hold between a mere bystander and the happening. If we accept that agent-regret is a form of guilt, then we can say that what distinguishes guilt from emotions like mere spectator-regret or dismay is that the *self* figures in the former: the feeling of guilt takes as its target the self, or part of it. The account of guilt that we offer has to comply with this constraint.

To motivate the account of guilt that I'll propose, I want to look at an explanation of the phenomenon of agent-regret latent in the remarks that Williams makes in “Moral Luck.” Given that agent-regret is a form of guilt, guilt about what we are agentially responsible for, the hope is that we could extend this explanation of agent-regret into an account of guilt more generally.

Why, then, do we feel agent-regret? Why do we feel bad about ourselves when some bad arises as the perhaps unintended and unforeseeable consequence of our actions? (Why do our self-directed emotions operate on the basis of strict, rather than fault, liability?) The explanation that Williams hints at has two components. First is the claim that we *identify* with our agency: we treat our ability to act and to affect the world through our actions as one of our fundamental features. As a result of identifying with our agency, we experience self-directed

---

<sup>36</sup>“Moral Luck,” 28.

attitudes on the basis of the things that flow from our agency. Williams writes, for example, of our “identity and character as agents,” and that “one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for.”<sup>37</sup>

Second, Williams notes that our agency extends to all effects of our actions, and not just the intended or foreseeable effects. After all, our agency is our capacity to *act*, so it is involved in all of our actions, even to the extent that they are unintentional: the truck driver’s agency attaches to his killing the child (and so to the death of the child, the upshot of the action) just as it attaches to his driving down that road. Perhaps what we do intentionally is most centrally inside our agency, but what we do accidentally or through ignorance still falls in its the periphery.<sup>38</sup> So Williams writes that it a “large falsehood” that we could view ourselves as agents and detach ourselves from the unintentional effects of our actions, and that “one’s history as an agent is a web in which anything that is the product of the will is surrounded and help up and partly formed by things that are not.”<sup>39</sup>

Putting these two claims together yields the conclusion that we feel bad about ourselves for the bad effects of our actions, regardless of whether they were intentional or foreseeable, because some feature or capacity with which we identify—our agency—is involved in them. Because we our agency forms part of the core of how we think of ourselves, and our agency attaches to everything that we do, we identify with all of our doings, so feel bad about ourselves whenever something for which we are agentially responsible is bad. Because the truck driver cannot help but think of himself in terms of what he has done, and knows that killing the child (however unintentionally) is among the things he has done, his identity is bound up with the death of the child, and he feels terrible about himself as a result.

## 7 A new account of guilt

This explanation of agent-regret (guilt about agential responsibility) naturally suggests an account of guilt more generally, which I’ll call the *self-implication* account of guilt:

Guilt represents some part of the self as being implicated in a bad

---

<sup>37</sup>“Moral Luck,” 29f.

<sup>38</sup>In the same vein, David Enoch writes that the accidental consequences of our actions lie in “our penumbral agency” (“Being Responsible”).

<sup>39</sup>“Moral Luck,” 29.

state-of-affairs.

This account implies that guilt is *fitting* when some part of the self really is implicated in a bad state-of-affairs, and that we *feel* guilt when we take some feature that we regard as part of the self to be implicated in what we take to be bad. In the next section, I'll show how the self-implication account explains both guilt about wrongdoing and guilt about various forms of causal responsibility that are not necessarily forms of moral responsibility. In this section, I'll spell out the basic ideas in the account.

First, I'll elaborate on what I mean by "the self." To regard some aspect or feature as part of the self, in my sense, is to *identify* with that feature: to have that feature figure into a description of oneself that is psychologically entrenched, or form part of the core of how one thinks of oneself. I might, for example, regard my role as a parent or philosopher as part of the self, finding it difficult to conceive of myself except in terms of these roles. The opposite of identifying with some feature one has is to be detached from it: I might be detached from facts about my family history or physical features, recognizing them but not conceiving of myself in terms of them.

Note that this is an expressivist rather than descriptivist account of the self: it tells us not what it is for some feature to be part of the self, but rather what it is to *regard* some feature as part of the self. (Consider the analogy with moral expressivism: an expressivist account of moral judgment tells us what it is for someone to *think* that an act is morally wrong, but not what it is for an act to be morally wrong.) As such, it does not specify conditions on some feature's being part of the self, so is neutral on when a feature really is part of the self, and consequently on the fittingness of various kinds of guilt.

I am using "identification" in a similar way to how Harry Frankfurt uses the term.<sup>40</sup> To identify with a desire, according to Frankfurt, is to take that desire to be an expression of who one really is; in identifying with a desire, I form a higher-order desire that I be motivated by that desire. In contrast, I might withdraw from certain desires that I have, like addictive ones, taking them not to express my self and desiring that they not motivate me. Similarly, consider what Christine Korsgaard calls a "practical identity": as she puts it, "a description under which you find your life to be worth living and your actions to be worth undertaking."<sup>41</sup> And Meir Dan-Cohen writes of "the self's ability to *identify*

---

<sup>40</sup>"Identification and Externality," in *The Importance of what We Care about* (Cambridge University Press, 1998).

<sup>41</sup>*The Sources of Normativity* (Cambridge University Press, 1996), 101.

with various elements and thereby integrate them into itself, or to distance itself from them by objectifying them and holding them at arm's length."<sup>42</sup>

In addition to more mundane, concrete objects of identification like social role or profession, one can also identify with more abstract features. Among these more abstract features are capacities like volition or will, the capacity to form decisions about what to do. After all, our ability to make practical decisions forms part of the core of how we think of ourselves. Within philosophy, for example, Kant's moral philosophy and classical Stoicism focus exclusively on the will. In these systems, we are essentially embodied wills, and other features that we have are treated as being basically outside the self. (Recall that Kant thought that acting on one's desires is heteronomous, in the same category as being compelled by others, thereby treating even our desires as something external.) This is the sense in which Williams takes Kantianism to be attractive (as he puts it, "only superficially repulsive"<sup>43</sup>), since it places what it takes to be most important about us, the will, entirely within our control.

Second, many of the capacities that we identify with function to *produce* certain kinds of objects. The objects that a capacity functions to produce will be the expressions or manifestations of that capacity in the world. Our volition, for example, is the capacity to produce intentions, and our intentions manifest our volition. Now, our volition may also indirectly produce (in the sense of causally generate) other objects, production of which is not its function: because it produces intentions, which generate actions, which create changes in the world, our volition indirectly produces those changes, even to the extent that they are unintended. The death of the child, for example, is attributable ultimately in some sense to the truck driver's volition, since the truck driver's decision to drive down that road led to the child's death. But there is a sense in which the unintended effects of our actions do not flow directly from our volition, while our intentions do. When a capacity produces certain objects of a kind production of which is its function, in the way that our volition produces intentions, I'll say that it is *implicated in* the existence of those objects.

Each of our productive capacities, then, specifies a range of objects to which we are connected through our having that capacity. Phenomenologically, our identification with certain capacities allows the self to extend into the world, marking the products of those capacities as "ours."<sup>44</sup> If we identify with a given

---

<sup>42</sup>"Responsibility and the Boundaries of the Self," *Harvard Law Review* 105, no. 5 (1992): 966.

<sup>43</sup>"Moral Luck," 21.

<sup>44</sup>Consider what William James writes about the extended self:

capacity, we relate to what the capacity is implicated in first-personally, treating the quality of the objects produced by that capacity as grounds for having various reactive attitudes toward ourselves. Because, for example, we identify with our volition, treating it as part of the self, we experience a sense of ownership over our intentions, since our volition is implicated in our intentions. To the extent that these intentions are good or bad, we might feel good or bad about ourselves as a result. On the self-implication account, guilt is one such self-reactive attitude: again, we feel guilt when we take a capacity that we regard as part of the self to be implicated in a bad state-of-affairs.

## 8 Explaining different forms of guilt

We are now in a position to show how the self-implication account of guilt explains both guilt about wrongdoing and guilt about being merely causally responsible for what is bad.

I'll take the case of guilt about moral responsibility first. On my account, the fact that we identify with our volition explains why we feel guilt about wrongdoing. After all, wrongdoing requires some defect in our volition: we do wrong either by forming a bad intention, or (in the case of negligence) by failing to intend the necessary precautions against some foreseeable bad. If I intentionally harm an innocent person, I do wrong by forming a bad intention to bring about a bad state-of-affairs. Similarly, if I drive recklessly, I do wrong by failing to intend to act with the safety of others in mind. In both these cases, there is a bad state-of-affair constituted either by our having formed a bad intention, or by our having failed to form a sufficiently good intention. Because our volition is implicated in this state-of-affairs, we feel guilt when we take ourselves to have satisfied the conditions on wrongdoing.

Our new account of guilt thus naturally explains for the familiar cases of guilt, those in which someone takes herself to have done something morally wrong. If our volition were the *only* capacity with which we could identify, then our account would predict that these are the only cases of guilt possible. So we

---

In its widest possible sense . . . a man's Self is the sum total of all that he *can* call his, not only his body and his psychic powers, but his clothes and his house, his wife and children, his ancestors and friends, his reputation and works, his lands and horses, and yacht and bank-account. All these things give him the same emotions. If they wax and prosper, he feels triumphant; if they dwindle and die away, he feels cast down—not necessarily in the same degree for each thing, but in much the same way for all (*The Principles of Psychology*, vol. 1, ch. 10).



might think of the received account of guilt as implicitly endorsing a conception of the self on which we are essentially our wills, to the exclusion of any other capacities: if we are essentially our wills, then the only important part of us that can be implicated in what is bad is our will.

But for most of us, our identification extends past our volition, and this is what allows for forms of guilt without perceived wrongdoing. Again, as Williams pointed out, most people identify with their agency, what they do, in addition to their volition, what they will. (To most of us, identifying solely with our volition, hence excluding our agency from the scope of the self, might seem like an “insane” form of auto-alienation.) As a result of identifying with our agency, we relate first-personally to our actions, even to the extent that those actions are unintentional. In other words, because we cannot help but think of ourselves as agents, our identity is bound up with what is our doing, even to the extent that we are not morally responsible for what is our doing. This, again, is what allows for the possibility of guilt about mere agential responsibility that does not rise to the level of moral responsibility.

So far, I’ve shown how the account explains guilt about moral wrongdoing, as well as about agential responsibility, but not how it explains guilt about causal responsibility more generally; we have not seen how it explains, for example, guilt about instrumental or indirect causal responsibility. I claim that many people identify with, in addition to their volition and agency, their *causal presence*: that is, their capacity to stand in causal relations to other things, to affect and be affected by other parts of the world. After all, we do think of ourselves as people *because of whom* things happen, not just through what we will or what we do, but more generally: through our bodily processes, through our experiences, through what others do because of us, and through other things that are not always within our control.

As a result of identifying with our causality, we relate first-personally to the things in which our causality is implicated, that is, to the things that we have causal involvement in. We feel a sense of ownership over the events or states-of-affairs that we take ourselves to be causally responsible for, since we take them to exist because of us, and we identify with our ability to bring things about. When the outcome is good enough, we might even experience pride about our merely causal role in relation to it. Consider, for example, Alexander Fleming’s serendipitous discovery of penicillin, for which he (and two others) were awarded a Nobel Prize in medicine: while Fleming was researching bacteria, his habitual untidiness caused a petri dish to be contaminated with a mold that destroyed bacteria around it. It seems that Fleming never realized the medical potential of

the mold, and abandoned research on it after some unpromising experiments. It took other scientists to realize that penicillin could be used as a miracle drug, to purify it, and to establish production chains; Fleming, then, deserves little credit for the development of the drug aside from its accidental discovery. Nonetheless, it would have been completely intelligible for him to feel pride about the great good that the development of the drug brought about, even though his role in that was almost entirely a matter of dumb luck. Cases of accidentally causing some smaller good can also be occasions for pride: A couple whose wedding is the precipitant for two of their friends to meet each other and fall in love might be proud, many years later, for having brought that outcome about. The victim of a crime might be proud about the passage of much-needed legislation spurred on by the crime, simply because he knows that the legislation exists because of his experience.<sup>45</sup>

That we identify with our causal presence is what, according to my account, explains guilt about the other forms of causal responsibility we discussed, instrumental and indirect. Consider the examples we used to illustrate guilt about these forms of causal responsibility, *Elvis* and *rescuer*. In both these examples, the subjects take their causal presence to be implicated in some bad state-of-affairs: Elvis thought that his twin brother died because of him, and the man in need of rescue thinks that the firefighter died because of him. Since they identify with that presence, they take a part of the self to be implicated in what is bad, and they experience guilt about what they take themselves to be causally responsible for.

Note that, in addition to making predictions about when a subject feels guilt, the self-implication account also makes predictions about the *intensity* of guilt that a subject feels, and when a subject *fails* to feel guilt at all. After all, the account specifies three dimensions variation along which determines variation in

---

<sup>45</sup>This is simplifying things a bit. Psychologists sometimes distinguish between two forms of pride, *achievement-oriented pride* and *hubris* (used as a technical, non-evaluative term). (See, for example, Jessica Tracy and Richard Robins, "Putting the Self into Self-Conscious Emotions: A Theoretical Model," *Psychological Inquiry* 15, no. 2 (2004): 103–125.) Achievement-oriented pride focuses on the outcomes that one is responsible for, whereas hubris focuses on the traits that one has: so I might feel achievement-oriented pride about having helped someone, and hubris about being a generous or kind person. Hubris, then, seems to be the positive counterpart to the emotion of shame: hubris represents the subject as having desirable features, just as shame represents her as having undesirable features. In contrast, achievement-oriented pride is the positive counterpart to guilt: it represents the subject as responsible for a positive outcome, just as guilt represents her as responsible for a negative outcome. When I use "pride" in the examples above, I'm using it in the sense of achievement-oriented pride rather than hubris.

the intensity of guilt: the degree of identification with a given capacity, the degree of perceived implication of that capacity in the state-of-affairs, and the degree of perceived badness of the state-of-affairs. A difference between two otherwise identical subjects in any of these dimensions would create a difference in the intensity of the guilt that they feel. Similarly, failing to identify with the capacity implicated in the state-of-affairs, failing to take that capacity to be implicated in the state-of-affairs, and failing to take the state-of-affairs to be bad would result in a subject's failing to feel guilt. Just as the received account would predict that a subject does not experience guilt over acts that he does not take to be morally wrong, the self-implication account has the tools to explain the absence of guilt in someone who seems like a fitting subject of guilt.

Finally, note that the self-implication account of guilt is not committed to the *fittingness* of cases of guilt without actual wrongdoing, even though it shows how they might be fitting. After all, the account implies that a case of guilt is fitting if and only if some part of the subject's self has been implicated in some bad outcome, but it does not specify which features really are parts of the self. If a Kantian or Stoic conception of the self, according to which the will alone lies within the boundaries of the self, is correct, then only cases of guilt about actual wrongdoing are fitting. We will still be able to explain, however, the *possibility* of guilt without actual wrongdoing, in terms of the subject's mistakenly regarding some feature (like his agency or causal presence) as part of the self.

## Conclusion

I'll conclude by showing that the self-implication account of guilt satisfies the desiderata that we posed.

First, we've seen that the account explains the familiar cases of guilt, guilt about bad outcomes that we are morally responsible for: when we are morally responsible for a harm to another, our volition is implicated in that outcome, and we identify with our volition. Second, it explains guilt about mere causal responsibility in terms of our identifying with our agency (in the case of agential responsibility) or our causality altogether (in the case of instrumental or indirect responsibility). The account also specifies in virtue of what some feel guilt about what they are merely causally responsible for while others do not: this difference stems from a difference in the capacities with which they identify. Those who feel guilt about causal responsibility identify with their causal presence, treating their ability to stand in causal relations as part of who they are fundamentally, whereas those who feel guilt only about wrongdoing identify solely with their

volition, treating their ability to will as the sole constituent of who they are fundamentally and their causality as incidental. Third, the account is stringent enough not to count spectator-regret, the distress that a mere spectator to a tragic event feels, as a form of guilt. After all, in the case of spectator-regret, the subject is not connected to the unfortunate outcome through any of the productive capacities that she could identify with; the event cannot be attributed to her in any sense. Fourth, the account explains these cases of guilt in a way that captures the self-directed nature of guilt: according to it, guilt represents a part of the self as bound up with the bad outcome.

Finally, the account is open ended enough to allow for guilt in other kinds of cases. After all, the account predicts that we experience guilt whenever some feature with which we identify is implicated in some bad, and it leaves the question of which features we identify with open. Specifying the right features could explain other kinds of guilt without wrongdoing, like vicarious guilt or survivor guilt. For example, in the case of vicarious guilt, even if we are individually innocent for some bad outcome, identifying with the volition of the people out of whose actions the outcome *does* flow could make us relate to that outcome first-personally, thinking of it as “ours,” which might cause guilt. And similarly, in the case of survivor guilt, perhaps we identify with our capacity to stand in relations to others, regarding that capacity as part of the self. To the extent that these relations are ones of undeserved inequality (as when we survive something that others do not), we might think that they are to that extent bad, and our relational capacity becomes implicated in the existence of something bad. This is obviously very sketchy, but it shows how the account presented here might explain these cases of guilt.

The received account has to say that subjects of these forms of guilt are making a mistake, construing themselves as having done wrong when they haven't. But if the self-implication account is correct, the subjects may not be making any mistake; these forms of guilt without perceived wrongdoing, like guilt about mere causal responsibility, may be fitting responses to situations that the subjects had no control over.