The Manifestation Account of Responsibility

Adam Lovett and Stefan Riedener

Abstract. According to self-expression views of responsibility, we're responsible for something insofar as it expresses who we are as persons. We think this general view is correct. But existing versions of it face a serious problem. They focus entirely on our responsibility for actions or mental states. Yet we take it that we can also be responsible for things out in the world—the fact that someone dies or survives, for example. Existing self-expression theories cannot account for this. So they cannot explain the full range of things for which we're responsible. To solve this problem, we propose the manifestation account of responsibility. On this account, you're responsible for something to the extent that it manifests you as a person. This view explains our responsibility for actions and mental states as well as external things. It can be defended against the most serious objections. And as an extra payoff, it also explains the gradability of responsibility. So, we argue, the manifestation account of responsibility is worth taking seriously indeed.

1. Introduction

Colin insults Delilah. He says she sounds like a hen when she's angry. Delilah gets furious, punches Colin on the nose, and breaks it. Both are normal adults, and neither has been manipulated, intoxicated, brainwashed or anything. Intuitively, they're morally responsible for what they're doing. But why exactly are they?

Various answers have been proposed. Some think Colin and Delilah are responsible because they stood outside deterministic causal chains: perhaps they acted as agent-causes (Chisholm 1982), or otherwise initiated causal chains themselves (Kane 1996). Some think they're responsible because they acted with a certain kind of ability or control: perhaps they were able to act without compulsion or constraint (Schlick 1939; Ayer 1954), or to respond properly to their reasons (Wolf 1993; Fischer and Ravizza 1998; Nelkin 2011).

But according to one prominent story, Colin and Delilah are responsible because their actions express something about themselves. Perhaps they express their "deep selves"—the dispositions that make up who they are

at their core (Frankfurt 1971; Watson 1975; Wolf 2003; Sripada 2016; Gorman 2022). Or perhaps they expresses their general "quality of will" (Strawson 1962; Wallace 1996; Arpaly 2002; Arpaly and Schroeder 2014) or reason-sensitive attitudes (Scanlon 1998; 2008; Smith 2005; 2008; Reis-Dennis 2018). Either way, the idea is that when our actions somehow express ourselves, we're responsible for them. Call this a *self-expression account* of moral responsibility.

We think that self-expression accounts are basically correct. In fact, we'll just assume this for the purposes of the paper. We'll assume that we are responsible for certain things (pace skeptics like Pereboom 2001; 2014; Strawson 1986; 1994; Levy 2011; or Waller 2011), and that that's roughly because these things somehow express who we are. However, we think existing self-expression theories face an important problem. In our everyday practice, we hold Delilah responsible not just for punching Colin or intending to punch him. We also hold her responsible for the fact that his nose is broken. More generally, according to our common practice, we're responsible not just for our actions or mental states, but also for things out in the world. Yet existing self-expression accounts focus entirely on our responsibility for actions or mental states. They tell us when an action or a mental state expresses ourselves. But they don't say how some things in the outer world can do so. And what they say about our responsibility for actions and mental states doesn't generalize to our responsibility for external things. So existing self-expression accounts cannot properly account for the full range of things for which it seems that we are responsible.

In this paper, we propose a solution to this problem. Our core idea is to employ the concept of manifestation. We'll suggest that you're responsible for something to the extent that that thing *manifests* you as a person. The notion of manifestation has been used in self-expression accounts before (Dewey 1891: 160; Arpaly and Schroeder 2014: 170). But no one has analyzed the notion in detail, or explored what work it can do. We'll show how a carefully explicated notion of manifestation pinpoints the relationship between yourself and the things you're responsible for. Here's the plan. In the next section we'll motivate the idea that we're responsible for things besides actions and mental states. In section 3, we'll see that existing self-expression accounts struggle to capture this. In section 4, we'll argue that our manifestation account of responsibility solves this problem. In section 5, we'll address some objections to our account. And in section 6, we'll argue that the account has another theoretical payoff. It not only captures what kinds of things we're responsible for. It also explains the *degree* to which we're responsible for them.

First, though, let's clarify what we mean by "responsibility." As has been common since Strawson (1962), we think for you to be responsible for something is for it to be appropriate to have certain reactive attitudes towards you concerning that thing. For example, Colin and Delilah are both blameworthy for their actions. It is appropriate to feel some indignation at Colin for insulting Delilah and at Delilah for hitting Colin. Thus they're responsible for these actions. The relevant attitudes here must be understood broadly. They're not just attitudes of moral praise and blame. You can be responsible for morally neutral things—like wearing yellow pajamas.[1] And you can be responsible for morally bad things, but have an excuse—like giving someone a glass of poison you justifiably, but falsely, believed to be water. In these cases, it may be inappropriate to blame or praise you for your actions. But it's still appropriate to have a morally neutral reactive attitude to you: perhaps to silently acknowledge what you did, to think it reflects certain things about you which are morally neither here nor there, to be curious or surprised or indifferent about why you did it and so on. These responses are still importantly different from the attitudes we take towards non-agential things like earthquakes, mathematical truths or the color of your eyes. Thus for you to be responsible is for it to be appropriate to have moral reactive attitudes towards you in this broad sense.[2]

2. What Are We Responsible for?

Any account of moral responsibility needs to identify what kinds of things we're responsible for. The starting point of this paper is that we're not just responsible for our actions or mental states. We're also responsible for *external facts*—i.e., states of affairs that are neither identical to nor constituted by our actions or mental states. We've given an example of this already. Delilah is responsible not just for her anger or her punch, but also for the fact that Colin's nose is broken. There are innumerable examples of this. In 2012, captain Francesco Schettino drove the *Costa Concordia* cruise ship into a rock. He had neglected various safety measures both before and after the crash, and thus caused 33 deaths. Schettino is responsible for how he steered his ship (an action), and for his recklessness (an attitude). But he's also responsible for the fact that 33 people died. Or consider Irena Sendler, who saved numerous children from the Warsaw

[1] For a similar point, see Talbert (2025: 4).

[2] On this understanding, for you to be responsible *just is* for such attitudes to be appropriate. One might take a slightly different view, on which your responsibility *grounds* the appropriateness of these attitudes. Not much hinges on this difference for us. See e.g. Menges (2020) for discussion.

Ghetto during World War II. She smuggled them out and gave them shelter. Sendler is responsible not just for her actions and attitudes, but also for the fact that these children survived the war.

We take these to be intuitive judgments. Or anyway, we take them to be reflected in our common practice of praise and blame. We praise and blame people *for* certain things. These objects are (part of) what makes our attitudes fitting. And they (in part) constitute the form that these attitudes take: to blame someone for something is to hold *that thing* against them; to praise them for something is to celebrate them for *this thing*. And that we can appropriately praise and blame people for something means they're responsible for it. Now one common reaction to Delilah would be to blame her for hitting Colin: to hold her action against her. Yet another equally common reaction, we think, would be to blame her for the fact that Colin's nose is broken: to hold this external fact against her. We would typically confront her not just by saying "you punched him!", but also by exclaiming "his nose is broken!". Our attitudes towards her would usually be colored not just by the fact that she moved her arm, but also by his injury. They would typically be very different from how they would be if Colin wasn't hurt at all. But this just means that according to our common practice Delilah is responsible for Colin's broken nose.

Now we personally are inclined to take this very seriously. It is a serious drawback for an account of responsibility, we think, if it cannot vindicate our common practice; if it cannot explain that we're responsible for external facts. However, one might deny this. The most principled reason for doing so is an anti-luck intuition (à la Nagel 1979). One might think that luck—or the influence of factors beyond our control—always defeats moral responsibility. It was partly due to factors beyond Delilah's control that Colin's nose is broken. So one might think she isn't responsible for his injury. More generally, one might think external facts are never entirely within our control, or always partly a matter of luck. So one might think that our practice is misguided, and that we're never responsible for such facts.

The question of moral luck is complex. For what it's worth, we personally think the idea that luck defeats responsibility leads to skeptical conclusions. That's because our actions and attitudes are also subject to luck. When Delilah intended to hit Colin, her arm might have cramped, or she might have fainted. And before she even managed to form any anger, she might have been distracted by a plane crash near her. It's never entirely within our control to perform certain actions or have certain attitudes. So we think if any amount of luck defeats moral responsibility,

we aren't responsible for anything.[3] And we think such skepticism is implausible. Thus, we believe, luck cannot undermine responsibility.

However, perhaps this argument can be resisted. Perhaps there's a way of saying that luck undermines responsibility for external facts, but not for attitudes and actions.[4] We don't intend to settle this here. So, we'll proceed on a conditional basis. If we can reject our common practice, and the argument we've just sketched, perhaps we don't need an account of our responsibility for external things. But *if* we ought to take our common practice seriously, or if that argument is sound, we need an account of responsibility that can explain how we're responsible for external facts. And it is on this assumption that our paper proceeds. We think this assumption is sufficiently plausible and widespread to be very much worth exploring. So let's now see whether existing self-expression accounts can capture our (assumed) responsibility for external facts.

3. Self-Expression Accounts of Responsibility

Self-expression accounts say you're responsible for something if and only if it somehow expresses your self. There are two key questions such a view must answer. First, what is your "self"? That is, what is it about you that must be expressed in something such that you're responsible for that thing? Some people are relatively restrictive on this: they think you're only responsible for expressions of your "second-order volitions" (Frankfurt 1971), your actual "values" (Watson 1975) or "cares" (Sripada 2016) or the attitudes you "approve of" (Gorman 2022). Others are more permissive: they suggest you're responsible for everything that expresses any of your desires (Arpaly and Schroeder 2014) or judgment-sensitive attitudes (Scanlon 1998; 2008; Smith 2005; 2008; Reis-Dennis 2018). The general question is: what about you generates responsibility, and what doesn't?

Second, what is "expression"? That is, what is the relationship you must bear to something to be responsible for that thing? The relationship is almost certainly a causal relationship. But not any causal relationship will do. We're not responsible for all the causal effects of our volitions, values or attitudes. Suppose on the way to the doctor to fix his broken nose, Colin meets a woman. They chat and soon fall in love. Ten years later they

---

[3] This kind of argument is given e.g. in Nagel (1979) (though he draws a somewhat different conclusion); a similar argument (to the effect that "luck is luck", or that a rejection of one kind of luck should plausibly imply the rejection of other forms) is given in Moore (1997: ch. 5) and Hartman (2017).

[4] For views along these lines, see e.g. Khoury (2018) or Talbert (2025).

have a daughter who another twenty years later steals someone's bicycle. This bicycle robbery is a causal consequence of Delilah's actions. Without them, Colin's daughter would never have come to be. But Delilah isn't responsible for this robbery. It doesn't *express* her in the relevant sense. So the general question is: which causal relationships generate responsibility, and which don't?

We are, for the most part, neutral on the first question in this paper (although we'll make a brief suggestion about it at the end). Our main focus is on the second question.

The problem for existing self-expression views is that they struggle to capture our responsibility for things out in the world. That is because they invariably focus just on responsibility for our actions or mental states. Gary Watson (1996), for example, says that "actions which "express" ourselves in the required sense are free actions" (Watson 1996: 227). The focus is solely on what actions we're responsible for. Tim Scanlon similarly says that "to claim that a person is blameworthy for an action is to claim that the action shows something about the agent's attitudes towards others" (Scanlon 1998: 128). Angela Smith extends the object of responsibility to attitudes. She thinks "an agent's responsibility for her actions and attitudes [is grounded] in the fact […] that they express *who she is* as a moral agent" (Smith 2008: 368). None of these authors give an account of what it is to be responsible for a thing out in the world.

So these authors don't give a full account of responsibility. Can we extend their stories to an account of responsibility for external facts? Unfortunately, we cannot. That is because these three authors do very little to answer the second question for self-expression theories. They don't provide a clear account of "expression", or of the relationship you must bear to an action or attitude for you to be responsible for it. And so we cannot say whether or when you bear that relation to external facts. Most self-expression theorists don't give us anything close to the materials to explain how you're responsible for things out in the world.

But some self-expression theorists do explicitly offer an account of "expression". They say *which* causal connections ground responsibility—or at least, responsibility for actions. Here are the two most prominent proposals today. For Nomy Arpaly and Timothy Schroeder (2014), expression is about rationalization: roughly, our actions express ourselves when our attitudes (not just cause, but) *rationalize* these actions. More precisely, we sometimes intrinsically desire good things—people's happiness, say. But we're sometimes indifferent to goods. And we sometimes intrinsically desire bad things—people's pains, for instance. These desires rationalize certain actions: they give us reasons, or make it

rational for us, to do these actions. And they can cause us to do these actions precisely in virtue of the fact that they give us such reasons. Delilah might want to hurt Colin, and this desire might cause her to punch him in virtue of the fact that it gives her a reason to do so. According to Arpaly and Schroeder, we're praiseworthy for actions that are caused in this rationalizing way by our intrinsic desire for goods. And we're blameworthy for actions that are caused in this way by our intrinsic desires for bads or our indifference towards goods. The relationship of expression is essentially that of causation-in-virtue-of-rationalization (Arpaly and Schroeder 2014: 63–63, 170–71).

Relatedly, for Chandra Sripada (2016), expression is fundamentally about motivation: roughly, our actions express ourselves when our attitudes (not just cause, but) *motivate us* to do these actions. To explicate this idea, Sripada assumes that our minds are populated with different "action-directed psychological mechanisms" (Sripada 2016: 1216). These are the characteristic mechanisms by which attitudes yield actions—things like deliberation about our motives, doing what's associated with pleasant emotional experience, or reinforcement learning. Our attitudes exert "motivational influence" on our actions, or motivate us to do them, if they cause our actions in some such way. For instance, Delilah might care deeply about letting wrongdoers feel her disapproval. This might have associated the action of punching Colin with a positive sensation for her. And this in turn might have led her to punch him. According to Sripada, we're responsible for actions that are caused in some such motivating way by our relevant attitudes (for him: our "cares"). The relation of expression is that of causation-through-motivation.

Now these may be promising accounts of responsibility for actions—as the authors intended them to be. But as they stand, they simply don't apply to external facts. Our attitudes might give us a reason, or a motivation, to perform certain actions. Delilah's anger might give her a reason or motivation to punch Colin. But we cannot have a reason, or a motivation, for an external fact. Delilah cannot have a reason or motivation for the fact that Colin's nose is broken, say. Such facts are simply the wrong kind of thing to be so rationalized or motivated at all. So, given the letter of these views, they cannot explain our responsibility for external facts. Or at least, they cannot do so directly.

But perhaps they can do so indirectly. A natural line to take for these authors—or indeed, for any self-expression theorist—is to explain responsibility for external facts indirectly, via responsibility for actions. They might say we're responsible for an external effect of our actions if and only if we're responsible for the action, and the action caused the

effect in some specific manner. The most natural way to do this is via an epistemic criterion for expression. Self-expression theorists might say, roughly, that we're responsible for an external fact if and only if we're responsible for an action, and the action *foreseeably* caused that fact. Indeed, there's a large debate about such epistemic conditions, often outside self-expression theories (see e.g. Robichaud and Wieland 2017). So perhaps self-expression theories can just adopt the best proposal from that debate.

However, we're skeptical about this strategy. An initial worry is that such an account offers a disunified explanation for the objects of responsibility. It gives one (direct) story about responsibility for actions, and another (indirect) story about responsibility for external facts. This is not a devastating problem, but it is a theoretical drawback. Other things equal, it seems preferable to have a unified account of expression.

More importantly, we think there's just no convincing way of spelling this epistemic strategy out. What exactly should it mean that an external fact was a relevantly 'foreseeable' consequence of your actions? There are various proposals here. Some people adopt an undemanding criterion. On their view, an effect was foreseeable if you should have known that there was some nonzero probability that your action would cause this effect (Zimmerman 1986: 206). However, this renders the account too broad. Almost anything has *some* nonzero probability of happening as a result of your actions. Surely, Delilah could have known that it's *possible* that her punching Colin would lead to someone stealing a bicycle 30 years later. Even so, she isn't responsible for that bicycle robbery. The present account is too permissive.

Other people adopt a more demanding criterion. On their view, an effect was foreseeable if you should have believed (or should have had a credence above a certain threshold) that your action will cause this effect (Fischer and Ravizza 1998: 101; on one reading). However, this renders the account too narrow. We can be responsible for effects that were highly unlikely. Suppose Delilah places a bomb in Colin's house that explodes if and only if a certain ticket wins in the national lottery. It can be extremely unlikely that this action will kill Colin. Delilah shouldn't believe that it will do so (or have a credence above any remotely demanding threshold). But surely, if that ticket does win, and the bomb explodes, Delilah is responsible for Colin's death. So this account is too restrictive.

Still other people have adopted a more nuanced criterion. On their view, an effect was relevantly foreseeable if you could have known that your action increased the probability of that effect (Nottelmann 2007: 191; Nelkin and Rickless 2017: 120; Peels 2017: 177). Perhaps Delilah's punch didn't increase the probability of a bicycle robbery 30 years later. But her

placing of that bomb did increase the probability of Colin's death. So this view might look more promising. But we think it's also too narrow. And that's because we can be responsible for things that were overdetermined. Suppose Delilah knew that if she doesn't punch Colin, Mike Tyson will do so, and will certainly break Colin's nose. If out of her anger she punches him first, she doesn't increase the probability of him breaking his nose. Indeed, she may even lower it. Still, she's responsible for Colin's injury. Thus this account is overly restrictive as well.

No doubt there's more to be said.[5] But ultimately, we think we need a novel account of expression—an account that's not epistemic in nature, but metaphysical. So let's now outline our own proposal.

4. The Manifestation Account of Responsibility

Our suggestion is that "expression" should be cashed out as manifestation. The language of manifestation isn't new to discussions of responsibility. John Dewey, for example, long ago said:

> It is because a result flows from an agent's ideal for himself […] that the agent feels himself responsible. The result is simply an expression of himself; a manifestation of what he would have himself be. (Dewey 1891: 160)

More recently, Arpaly and Schroeder said that we're blameworthy for an action insofar as it "manifests ill will […] through being rationalized by it" (2014: 170). It's natural to read these philosophers as claiming that the responsibility-generating relationship is that of manifestation. But none of them take the language of manifestation very seriously. They use it as a tool to express a theory that could easily be expressed in different terms. And they don't provide any explication of "manifestation" itself. We think taking the language of manifestation more seriously solves the problem we've raised for self-expression views.

To show this, we first explain what we mean by manifestation. At its core, manifestation denotes the actualization of dispositions. When some disposition *D* gets actualized in some fact *a*, *a* manifests *D*. Consider the fragility of a vase. This is its disposition to shatter when dropped. Suppose the vase shatters, because it is fragile and because you dropped it. Then the shattering actualizes, or manifests, the fragility. Importantly however,

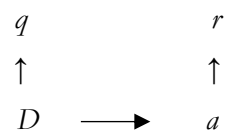
---

[5] One person who tries to find a solution to our problem, or a closely related one, is Montminy (2019; 2022). He ultimately concludes that no solution can be found (2022: 447). We think the manifestation account we're going to present resolves Montminy's problem.

not all causal consequences of the fragility are manifestations of it. Suppose in fetching the vase's shards from under the sofa, you find a long-lost earring. The fact that you found your earring is a consequence of the vase's fragility. But it isn't a manifestation of it. The vase's fragility is a disposition to shatter when dropped. It's not a disposition to make people find their jewelry. So only the shattering manifests the fragility.[6]

We take the notion of manifestation, or at any rate actualization, to be one about which we often have pre-theoretic judgments. Think about Rosa Parks refusing to give up her bus seat in 1955 Alabama. Her staying seated in the face of arrest was a manifestation, or actualization, of her courage. Likewise, consider Idi Amin torturing his political opposition. Their broken bodies were a manifestation of Amin's cruelty. Or contemplate Lionel Messi playing a brilliant game. Argentina's winning the match is a manifestation of his ability. In all these cases we have some natural disposition—courage, cruelty, footballing skill—and, when that disposition is actualized, the actualization is a manifestation of the disposition. But in none of these cases are all the causal consequences of the disposition manifestations of it. Messi's skill, for example, increased the number of people called 'Lionel', and thus the amount of butter bought by people with this name. But this increased namesake-butter-consumption is no actualization of Messi's skill. We take these examples to show that we have an intuitive grip on a notion of actualization. Actualization may be a species of causation but is not the same as it.

We've just said that actualization is the core notion of manifestation. But we think the concept of manifestation is often used in a slightly broader way, such that it encompasses grounding relations. Or anyway, that's how we will use it. Suppose disposition *D* grounds some fact *q*, and its actualization *a* grounds some fact *r*—as in the following diagram, where the horizontal arrow indicates actualization, and the vertical arrows indicate grounding:



[6] For a helpful overview of the metaphysics of dispositions, see e.g. Choi and Fara (2018). For some other attempts to use manifestation to analyze normatively interesting notions, see Turri (2011) and Mantel (2017).

As we understand the term "manifestation", both *a* and *r* manifest both *D* and *q*. As an example, consider again the fragility of our vase (*D*), and its actualization in the fact that it shatters (*a*). Now suppose the fragility of our vase partially grounds the fact that nothing lasts forever (*q*). Then, the fact that our vase shatters manifests (not just its fragility, but also) the fact that nothing lasts forever. Similarly, suppose the fact that our vase shatters partially grounds the fact that all vases in the room shatter (*r*). Then, it's (not just the shattering of our vase, but also) the fact that all vases shatter which manifests the fragility of our vase—or indeed, the fact that nothing lasts forever. Thus manifestation is defined in terms of both actualization and grounding: manifestation relations always involve actualization and sometimes flow through chains of ground. Or so at least we'll assume. We think this aligns with common uses of the term. But our everyday notion of "manifestation" is fuzzy. So we don't understand this notion of manifestation as the uniquely correct conceptual analysis of our everyday concept. It's one way—perhaps among others—to sharpen our notion in a theoretically useful manner.

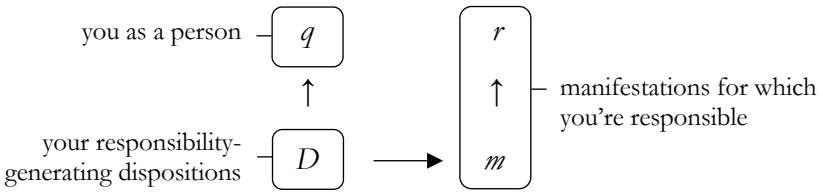
We now use this notion of manifestation to formulate an account of responsibility. Let's assume that you're responsible for what constitutes you as a person. Manifestation is how you become responsible for things outside of yourself. So we propose the

> **Manifestation account of responsibility:** You're responsible for *p* if and only if *p* constitutes or manifests you as a person.

This requires some further explanation. In particular, it raises the first question we've posed for self-expression theories: What constitutes "you as a person"—or, equivalently, what is the "self"? As we've indicated, we want to remain largely neutral on this question. You as a person might be constituted by your second-order volitions, or your values, or all your judgment-sensitive attitudes, or whatever. The manifestation account is compatible with different views here. It presupposes only that you are in part constituted by dispositions. But that is anodyne. The relevant features of you can all be thought of as being (or involving) dispositions. For you to desire to drink water is partly for you to be disposed to drink a glass of liquid when you believe it's water. To intend to eat is in part to be disposed to eat. To value others getting water is in part to be disposed to feel bad if they have no water. So the manifestation account is compatible with a wide range of views about which of your features must be expressed in something for you to be responsible for it—as long as they're dispositions. For the purposes of this paper, we'll thus take you as a person to be

constituted by the set of your "responsibility-generating dispositions"—leaving open what precisely these are.

Let's illustrate the resulting picture in a diagram again:



Thus, on the manifestation account of responsibility, you as a person are the totality of your responsibility-generating dispositions. And you're responsible for all of these dispositions, as well as for every fact that's an actualization of these dispositions, or that is grounded by such an actualization.[7] Now this is an account of responsibility, not of praise- or blameworthiness. As we said, you can be responsible for things you're not praise- or blameworthy for—such as wearing yellow pajamas. But the account can easily be extended to explain praise- and blameworthiness too. Roughly, we suggest, you're blameworthy for something if that thing manifests one of your morally bad dispositions. And you're praiseworthy for something if it manifests one of your morally good dispositions.[8]

How does this solve our initial problem? The key point is that our responsibility-generating dispositions are manifest not just in our actions and mental states, but also in facts out there in the world. Consider Delilah's irascibility. This disposes her to get angry at Colin and punch

---

[7] To emphasize, on the manifestation account, you're thus responsible not just for the manifestations of your dispositions, but also for the dispositions themselves. We think, in the sense of 'responsibility' we're concerned with, this is plausible: besides reproaching Delilah for particular irascible actions, intentions or feelings, we can reproach her for being an irascible person *tout court*. This doesn't mean that there are no differences between these loci of responsibility. Perhaps the fitting responses to people's actions or attitudes are different from those to people's dispositions: e.g., perhaps reactions to actions and attitudes have a kind of urgency (cf. Riedener 2020), or a certain retributive character (cf. Bennett 2002), which reactions to unactualized dispositions lack. Regardless, the idea that we're responsible for dispositions isn't central to our main argument in this paper. If one doesn't find it plausible, one can adopt a slightly revised version of the manifestation account which restricts responsibility just to manifestations.

[8] Note also that on this account, we're responsible for *facts*—rather than, say, events. We think this is an innocuous assumption. Events can just be understood as collections of facts.

him in the nose. So (assuming this irascibility is part of her responsibility-generating dispositions) these actions and mental states manifest her as a person, and she's responsible for them. But the irascibility also disposes her to injure him. After all, such injury isn't an accidental consequence of her getting angry. It's an actualization of her irascibility. The relation between Delilah's irascibility and Colin's broken nose is like that between Messi's brilliance and Argentina's winning the match—not like that between his skill and the butter consumption of Lionels. If she breaks his nose because of her irascibility, his broken nose manifests her as a person. And so, Delilah is responsible for the fact that Colin's nose is broken. Or consider Colin's rudeness. This not only disposes him to insult Delilah, but also to make her feel bad. Again, such feelings are a characteristic effect of such rudeness. And so when Colin insults Delilah, her hurt feelings manifest him as a person, and he's responsible for the fact that she feels hurt. So the manifestation account can cover responsibility for external facts—or in particular, for intentional harms.

It can also cover cases of negligence or recklessness. Take captain Schettino again. He was all too indifferent to the safety of his ship. This indifference not only disposed him to steer too close to the rocks. It also disposed him to have a fatal accident. And so the sunk ship and the resultant deaths are a manifestation of his indifference. It's not only positive attitudes like caring about, valuing, or desiring something that can be manifest in external facts. Attitudes of indifference or insensitivity are also routinely manifest in the world. So the manifestation account covers a wide range of cases.

Yet it avoids the problem of deviant causal chains that afflicts the simple causal account of expression. Consider again the bicycle robbery thirty years after Delilah's punch. This robbery is caused by Delilah's irascibility, but doesn't manifest it. Next-generation bicycle robberies are not a characteristic consequence of irascibility. When people are irascible, that doesn't typically lead other people to have children who then steal bikes. When this happens, it's a purely accidental consequence. So Delilah isn't responsible for that robbery. Manifestation is a causal relationship, but it is one that rules out accidentality.

How exactly does manifestation connect to notions like foreseeability? To begin with, manifestation itself is clearly not an epistemic notion. One shouldn't understand the connection between the fragility of a vase and its shattering in purely epistemic terms. It's also not the case that your dispositions are manifest in some outcome if and only if you could have known that your action had a nonzero probability of causing it, or you should have believed that your actions will cause it or increase its

probability. The bicycle robbery doesn't manifest Delilah, even if she could have known about its possibility. Yet if that low-probability-bomb kills Colin, then although she should not have believed that it will lead to his death, she is manifest in it. And she is manifest in Colin's broken nose even if she knew that her punch pre-empted a more devastating blow by Mike Tyson. More generally, we don't think there is any universal and straightforward connection between when your dispositions are manifest and your epistemic or doxastic states.

Nonetheless, facts about what we know or believe sometimes matter to moral responsibility on the manifestation account. For a start, such facts can affect the moral quality of our will (see e.g. Björnsson 2017). For example, often, the more confident you are that your actions will have bad effects, the more problematic the disposition is that will be manifest in that effect. Delilah's willingness to do something she thinks is very likely to harm Colin is morally worse than her willingness to do something she thinks has a tiny probability of doing so. So manifestations of the former attitude will be morally more problematic than manifestations of the latter and warrant more blame. Additionally, there are certain correlations between our epistemic or doxastic states and how much we're manifest in the effects of our actions. For instance, that we're ignorant of these effects may mean they don't manifest our attitudes at all. Suppose Delilah wants to kill Colin but has no idea that putting this powder into his coffee will do so: she reasonably thinks it is sugar rather than arsenic. Then, Colin's death will typically not manifest her desire to kill him. So manifestation cannot be understood in purely epistemic or doxastic terms, but still facts about what we know or believe can modulate the extent of our responsibility or blameworthiness.

Let's sum up. The core notion of our account is that of "manifestation"—which we've cashed out as a compound of actualization and grounding relations. We think this is sufficient to understand this notion. And we think the manifestation account of responsibility gets the correct verdict when it comes to responsibility for things out in the world. It explains our responsibility for mental states, actions *and* external facts with a single, unified notion. This, we think, is a significant theoretical advantage.

5. Objections

Let's address three objections to the manifestation account. First, one might worry that the account overgenerates. It makes one responsible for involuntary bodily reactions that, one might claim, one isn't responsible for. To bring this out, imagine that you're Donald Davidson's climber (1980: 79). You're on a climbing trip, your companion fell and is now

dangling off the wall on a rope. You just about manage to hold the rope. You want to survive and realize that this gives you a reason to let go of the rope. Unnerved by this, you start to tremble, unintentionally loosen your hold—and your companion plunges to their death. The trembling seems to manifest your desire to live. So on the manifestation account, you seem responsible for your trembling—and perhaps the resultant fall. Yet, one might think, you aren't responsible for either. Trembling wasn't an intentional action. It was precisely the result of a "deviant causal chain" that rules out responsibility. The more general phenomenon here is that our attitudes will dispose us not just to perform intentional actions but also to undergo some brute bodily changes—to tremble, blush, cry, and so on. The manifestation account says we're responsible for these things.

In response, we accept that in Davidson's case, you're responsible for your trembling, and perhaps even your companion's death. But we think you're not *blameworthy* for either of them. Again, you're blameworthy for something only if that thing manifests one of your morally problematic dispositions. Your trembling manifests your keen desire to survive, and perhaps your general lack of nerves. But these dispositions aren't morally bad. There's nothing morally defective about wanting to survive or being a little jittery. So, you're not blameworthy for their manifestations. Still, these manifestations are very different from, say, your height. There's no sense in which you're responsible for your height. Height isn't the right kind of thing to make any reactive attitudes towards you fitting. To reproach you for your height would be a category mistake. In contrast, to reproach you for trembling might rely on a faulty moral assumption—that jitteriness is morally bad—but it would not be a category mistake. We can be responsible for things like tremblings. They show something about who we are as persons. They make morally neutral attitudes appropriate—as does wearing a yellow pajama. So we wish to turn the tables around. We're not only responsible for intentional actions, but also sometimes for mere bodily movements (and indeed emotions). The fact that the manifestation account implies as much is a feature, not a bug.[9]

Let's turn to a second, opposite concern about the manifestation account. One might worry that the account undergenerates responsibility. The most salient cases here are those of strict liability. Suppose you own a cat. You're worried that it might get harmed or cause damage if it's outside, and thus always lock it up in your house. Yet one day a small meteorite shatters one of your windows, your cat escapes, and breaks a vase on the porch of your neighbor. One might say the breaking of the vase was a

[9] For other recent discussions of Davidson's case, see e.g. Arpaly and Schroeder (2014: ch. 3.4) or Wallace (2023).

freak accident, and so doesn't manifest any of your attitudes. So on the manifestation account, it seems, you're not responsible for it. However, one might think, you are responsible for it in some sense—for plausibly you ought to cover the costs. More generally, one might think, we're sometimes responsible for external facts that we cause, even if they don't reflect or express or manifest anything about us as persons.

To address this case, we distinguish two kinds of responsibility. We have given an account of moral responsibility—the relationship that makes reactive attitudes appropriate. But there is a second kind of responsibility—compensatory responsibility. This is the relationship to a harm that makes compensation appropriate. We think you're not morally responsible in cases of strict liability. It would be unfitting to blame you or be angry at you for your cat breaking the vase—or indeed to think it reflects anything (even morally neutral) on who you are as an agent. But you do owe compensation in such cases. This compensatory responsibility is what our judgments respond to. Of course, that requires that compensatory responsibility is genuinely distinct from moral responsibility. But many accounts of compensatory responsibility imply as much. Some people think compensatory responsibility arises from brute causality (Couto 2018), or faultless wrongdoing (MacCormick 1978), or is somehow relationship-based (Capes 2019). Or perhaps compensatory responsibility is largely conventional: what compensation you owe is determined by social norms. Such accounts all imply that you owe compensation in a broader range of cases than those in which you're morally responsible. So in cases of strict liability, we do have a kind of responsibility. But we lack moral responsibility.

Here's a third worry. One might challenge the way we've set up our problem. We've suggested we're not just responsible for our actions, but also for external facts. However, there's a thicker way of describing our actions, which includes external effects. Delilah did not just move her arm, or punch Colin. She also *broke Colin's nose*. Similarly, Schettino did not just steer here or there. He also *killed 33 people*. And Irena Sendler did not just move her body through Warsaw. She also *saved many people*. But breaking people's noses, or killing or saving them, are actions. So, one might deny that we're actually responsible for external facts. One might insist we're only responsible for (mental states and) actions—noting that some of our actions are 'thick' in this sense, and include relevant external facts in their descriptions. If that is so, the manifestation account is too broad. And maybe the existing accounts we've considered (like those by Arpaly and Schroeder and Sripada) are workable after all.

But we doubt they are. To begin with, according to our common practice, Delilah seems responsible not just for punching Colin or for breaking his nose, but also for the fact that his nose is broken. We can hold this external fact against her, just as we can do so with her actions or mental states. So the present line still seems somewhat revisionary to us. But moreover, even if we're only responsible for such thick actions, this doesn't change the dialectic much. For we still need to know *which* thick actions we're responsible for. Consider Schettino. He caused 33 deaths. And let's imagine he also caused a bicycle robbery, which took place ten years after the incident as a result of some accidental causal chain. Why is he responsible for the former but not the latter thick action? Arpaly and Schroeder or Sripada must say that his desires rationalized or motivated the former but not the latter. But why should that be? What exactly is the difference between these actions? No straightforward answer seems in sight. Or rather, as far as we see, the best answer to this kind of question must also proceed in terms of manifestation. His causing 33 deaths manifests him as a person, while his causing that bicycle robbery doesn't. And if that is true, we essentially end up with a version of the manifestation account, on which we're responsible for those actions and mental states that manifest us as a person. We'll proceed on the assumption that we are responsible for external facts—which, as we've said, we find more natural. But this alternative setup of the problem would actually not change much.

6. The Gradability of Responsibility

Our main argument for the manifestation account is that it explains the scope of responsibility: it explains why we're responsible for some things but not others. But another important feature of responsibility is that it comes in degrees. In this final section, we want to show that the manifestation account also illuminates this aspect of our phenomenon.

Intuitively, there are several ways in which responsibility comes in degrees. First, we're less responsible for things that have other causes, beyond our own actions or mental states. Consider Delilah hitting Colin. She's very responsible for his broken nose. Now suppose Colin was deeply in debt, due to some poor financial choices he'd made. The medical bills for treating his nose led him to lose his car. Delilah is responsible for Colin losing his car. But she's less responsible for this than for his broken nose. Second, we're less responsible for things that have other grounds, beyond the effects of our actions. Imagine Colin overall has a pretty bad life. This is partly due to the pain in his nose. But it's also due to the troubles he's having with his finances, his marriage and his career. Delilah is responsible

for the general fact that Colin's life is not going well. But her responsibility for that is far less than her responsibility for his broken nose. Third, we're less responsible for actions that are less characteristic of us. Suppose that Delilah genuinely opposes violence, generally tries to remain composed, and typically succeeds. But she's had a long day, she's bone-tired, and so in this case she responds with anger. Then she is responsible for punching Colin or for his broken nose. But she's less responsible than if her violence resonated with many aspects of her personality. In short, responsibility is scalar in a few ways. A good theory of responsibility should help explain what determines how responsible we are for something.

The manifestation account of responsibility does exactly that. The key point here is that the materials from which it's constructed—actualization and grounding—come in degrees. Let's see this first with actualization. Compare dropping a fragile vase on the floor with dropping a truck on that vase. In both cases, the vase shatters. But in the first case the fragility is more actualized by the shattering than in the second case. Generally speaking, it's natural to understand this in terms of causal contribution (Goldman 1999). Actualization is a specific causal relation, and causation comes in degrees. When you drop a vase, its fragility makes a large causal contribution to its shattering. When you drop a truck on the vase, the fragility makes a small causal contribution to the shattering. The vase would likely have shattered whether it was fragile or not. The extent to which something actualizes a disposition is proportional to the extent to which the disposition causally contributes to the thing.

Now consider grounding. We think it's pretty obvious that grounding comes in degrees. Some things make more of a grounding contribution to what they ground than other things. Think about the existence of the set of natural numbers up to and including 100. The existence of the numbers 1 to 99 together makes more of a grounding contribution to the existence of this set than does the existence of the number 100. Likewise, the existence of the United Kingdom is grounded more in that of England, Scotland and Northern Ireland together than in that of Wales alone. The existence of a conference is grounded more in the existence of the keynotes than the existence of a single session that a handful of people attend. Grounding comes in degrees. We're not going to give any general account here of what determines the extent to which one thing grounds another. But we think it's intuitively clear that there is such an extent.

It follows that there's a natural scalar notion of manifestation. Our notion of manifestation is a compound of an actualization relation and two grounding relations. So, we'll say, the greater the degree of these actualization and grounding relations, the greater the degree of

manifestation. And this generates a natural scalar notion of responsibility for us. We say you are responsible for *p to the extent that p* is a manifestation of you as a person. Suppose *p* manifests you as a person. Then you're more responsible for *p* the greater the degree of the actualization and grounding relations that link you to *p*. You're more responsible the more central *D* is to you as a person, the greater the degree of the relevant actualization, and the more central the actualization of *D* is to *p*.

This account of responsibility explains the intuitions with which we began this section. Consider, first, the causal consequences of our actions. When Colin loses his car, this might partly be an actualization of Delilah's intention to harm him. But, if it is, it's only a very weak actualization of it. A lot of things contributed to Colin losing his car—most saliently, his poor financial decisions—and so Delilah's intention only makes a small causal contribution to the loss. So Delilah is less responsible for the lost car than she is for the broken nose. Second, consider the constitutive consequences of our actions. A lot of things ground the fact that Colin's life is pretty bad: his financial precarity, his faltering marriage, his career frustrations. That he has a broken nose does make a grounding contribution to his life going badly, but a smaller one than it would make if his misery was entirely due to his nose. And so Delilah is a little responsible for Colin's poor quality of life, but much less responsible than she is for his broken nose. Third, consider how centrality to our character modulates responsibility. Intuitively, the more Delilah's irascibility is in character, the more central it will be to her as a person—or the more of a grounding contribution it will make to her self. And so on the above story, she'll be more responsible for an in-character punching than for an outburst of anger that's entirely out of character.[10]

Let's say more about this final idea. What determines how central some disposition is to you as a person? The manifestation account is compatible with many different answers to this question. But here's a story that suits its spirit well. Suppose the responsibility-generating dispositions include beliefs, desires, intentions and emotions. Such attitudes often aren't independent of each other. They're manifestations of one another, in that they help causally sustain one another. They form a kind of causal web. Imagine, for example, that you're a conspiracy theorist. You think 9/11 was an inside job. This belief sustains your other attitudes in that it non-

[10] To be clear, the manifestation account is compatible with denying that certain responsibility-generating dispositions are more central to who you are as person. It only commits us to the conditional claim that, if they are, then that affects responsibility. But we think some dispositions are more central to you than others. So we think this centrality makes a substantive contribution to responsibility.

accidentally causes them to persist. It might help sustain your belief that you can't trust the government, or your desire to avoid government surveillance. You might have and keep these other mental states, in part, because you think 9/11 was a conspiracy. If so, this belief is manifest in these other states. Now some of your relevant dispositions play a large role in sustaining your other mental states. They're central to the web of your attitudes. Other dispositions play very little role sustaining your other mental states. They're on the periphery of this web. Your relevant dispositions form a causal web the strands of which are connections of manifestation relations.

Our proposal is that, all else equal, you are more responsible for the manifestation of dispositions at the center of this web than at the periphery. This is how we should cash out notions of something being in- and out-of-character. When we say that Delilah's anger was out-of-character, we mean it is not very closely connected to her other mental features. Her anger doesn't play a big role in sustaining her beliefs about the world, or her desires and intentions. When we say that her anger was in character, we mean that her irascibility plays a big role sustaining her other mental features. More generally, we propose that that which is most central to who you are as a person is that which lies at the center of the causal web of your responsibility-generating dispositions.[11]

Let's note a few things about this proposal. To begin with, in some sense, this proposal is a "deep self" view of moral responsibility. It says what matters most in determining your responsibility are the attitudes and traits at the core of your personality. But it's more nuanced than most such views. Deep self views usually suggest that you're fully responsible for expressions of those attitudes that constitute your deep self, such as your second-order volitions (Frankfurt 1971), values (Watson 1975) or cares (Sripada 2016), or the attitudes you approve of (Gorman 2022), and not at all responsible for expressions of your other attitudes—the desires that you yourself don't endorse or prefer to get rid of, say. We think that's too crude. A good deep self view should acknowledge, as ours does, that depth is a matter of degree.

---

[11] Note that this is just an account of when a certain disposition is more or less central to you. Naturally, there's another (equally ground-theoretic) way in which something can be more or less a manifestation of you as a person—viz., if it expresses more or less of your dispositions. Suppose Delilah's punching Colin manifests her irascibility as well as her impatience. Other things equal, it will then be more a manifestation of her than if it doesn't manifest any of her attitudes towards time (in the way patience and impatience do).

This brings us to our second point. Given that we discount more peripheral dispositions, we're inclined to be permissive in defining the class of our responsibility-generating dispositions. We believe that *all* our normative dispositions are at least *somewhat* constitutive of who we are as people—even those we wish to get rid of or don't endorse. So we think you as a person are just the totality of your "reason-responsive dispositions." By these we mean your dispositions to (not) see certain facts as reasons for certain responses, to (not) respond accordingly to these facts, and to (not) cause certain characteristic results as a consequence of such responses. For you to have a certain character trait, desire, or value is for you to have a disposition of this form. For you to be irascible, say, is for you to be disposed to see various facts as reasons for getting angry, to get angry on this basis, and to then injure people, damage objects and so on. To desire water is to be disposed to see the fact that some action is a means to getting water as a reason to do that action, to actually do it on this basis, and to usually slake your thirst in such cases. Such reason-responsive dispositions contrast, say, with hayfever. Hayfever is a mere physical disposition to get a running nose or watery eyes when in contact with pollen. It's not a disposition to see anything as a reason for such responses. We think you are the collection of your reason-responsive dispositions. Some of these dispositions—like the ones you judge to be bad or wish to shed—are less central to you than your firmly endorsed attitudes. So you're less responsible for their manifestations. Still, they all play a role in who you are, and so you're somewhat responsible for the expressions of all of them.

To illustrate this account, suppose you're an unwilling heroin addict. You judge your addiction to be bad, want to get rid of it, and react with horror to what it drives you to do. Yet one day you steal a friend's money to feed your craving. Your addiction is a reason-responsive disposition: it disposes you to take the gratifications of heroin as an end. So on the view we've just sketched, you're somewhat responsible for your stealing. This is plausible. It would (pace Frankfurt 1971 or Watson 1975) not be unfitting for your friend to be somewhat irritated at you or demand an apology from you. You're still a person, not an object. However, to an important extent, your addiction is an isolated drive in your overall attitudinal web: it is not manifest in the bulk of your other attitudes. Your value-judgments, your overall plans for your life, your broader desires, concerns, and emotions don't manifest your desire for heroin. They're directly opposed to it. Thus your responsibility for its effects is less than it would

be if your stealing resonated with your entire person. And that is surely plausible too.[12]

We wish to emphasize that one doesn't have to accept this permissive picture of your "self" to accept the idea that expression is manifestation. One might think the responsibility-generating dispositions are just your second-order volitions, values or cares or whatever. Still, we think the picture we've sketched is a very plausible one—and this redounds to the plausibility of its components.

Let us end with a final, more speculative, point. Our argument may not only provide a consideration for our particular self-expression account over other such accounts. It may also provide a consideration for self-expression accounts in general over accounts of responsibility in terms of abilities or control. If what we've said is correct, all theories of responsibility need an account of how we can be responsible for external facts. Ability or control theorists have typically tried to account for this responsibility in indirect ways—"tracing" responsibility for external facts back to responsibility for appropriately abled or controlled actions (e.g. Fischer and Tognazzini 2009). Defenders of this tracing strategy typically either try to spell such traceability out in epistemic terms, or just don't say precisely when an external fact is traceable to an action. If our arguments are sound, very plausibly the best account of tracing is in terms of manifestation. But if that is so, there is theoretical pressure to explain all kinds of responsibility in terms of manifestation—not just responsibility for external facts, but also responsibility for actions and mental states. And that would mean to accept our account. But exploring the merits of this argument is a project for another occasion. So let us now conclude.

7. Conclusion

We've given an account of responsibility. You're responsible for something to the extent that it manifests your responsibility-generating dispositions. Thus you can be responsible—in just the same way—for actions and mental states as well as for external facts. This account of

[12] Note that other cases that are discussed in the literature may be different. Consider e.g. spasms due to epilepsy, or twitches or vocal tics due to Tourette syndrome (cf. e.g. Gorman 2022). Here people don't seem responsible at all. But in such cases, we think the underlying dispositions (epilepsy or Tourette syndrome) simply aren't reason-responsive dispositions in the first place. They're mere physical dispositions—like hayfever. And of course a willing addict—or at least one whose addiction has rearranged their values, plans, emotions and other attitudes—will be highly responsible for their actions. Their addiction is central to who they are.

moral responsibility connects to a much broader project. In other work, we've argued that the good life is the life that manifests or is manifested in the good (Lovett and Riedener 2024a). We've argued that the duties that arise in special relationships are generated by connections of manifestation (Lovett and Riedener 2025). And we've argued that obligations of gratitude, duties not to do harm, and duties not to intend harm, all arise from manifestation connections (Lovett and Riedener 2024b). Overall, we've advanced the idea that manifestation is important across moral philosophy. This idea corroborates the central claim of this paper—that manifestation is the connection that generates responsibility. The import of manifestation to responsibility is, we think, an instance of its import across various moral domains.

Acknowledgments. For very helpful comments on earlier versions of this article we thank Andreas Brekke Carlsson, Leonhard Menges, Jonathan Parry, Peter Schaber, Daniel Telech, Nikhil Venkatesh, participants of the Instituttseminar at the University of Bergen's Philosophy Department as well as two anonymous reviewers and an area editor at *Ergo*.

Bibliography

Ayer, Alfred J. (1954). Freedom and Necessity. In *Philosophical Essays* (271–284). MacMillan.

Arpaly, Nomy (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.

Arpaly, Nomy and Timothy Schroeder (2014). *In Praise of Desire*. Oxford University Press.

Bennett, Christopher (2002). The Varieties of Retributive Experience. *Philosophical Quarterly,* 52(207), 145–163.

Björnsson, Gunnar (2017). Explaining (Away) the Epistemic Condition on Moral Responsibility. In Philip Robichaud and Jan Wieland (Eds.), *Responsibility: The Epistemic Condition* (146–162). Oxford University Press.

Capes, Justin A. (2019). Strict Moral Liability. *Social Philosophy and Policy*, 36(1), 52–71.

Chisholm, Roderick (1982). Human Freedom and the Self. In Gary Watson (Ed.), *Free Will* (24–35). Oxford University Press.

Choi, Shungho and Michael Fara (2018). Dispositions. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

https://plato.stanford.edu/archives/spr2021/entries/dispositions/.
Couto, Alexandra (2018). The Beneficiary Pays Principle and Strict Liability: Exploring the Normative Significance of Causal Relations. *Philosophical Studies*, 175(9), 2169–2189.
Davidson, Donald (1980). Freedom to Act. In *Essays on Actions and Events* (63–82). Oxford University Press.
Dewey, John (1891). *Outlines of a Critical Theory of Ethics*. Register Publishing Company.
Fischer, John Martin and Mark Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.
Fischer, John Martin and Neal A. Tognazzini (2009). The Truth About Tracing. *Noûs*, 43(3), 531–556.
Frankfurt, Harry G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), 829–839.
——— (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5–20.
Goldman, Alvin I. (1999). Why Citizens Should Vote: A Causal Responsibility Approach. *Social Philosophy and Policy*, 16(2), 201–217.
Gorman, August (2022). Demystifying the Deep Self View. *Journal of Moral Philosophy*, 19(4), 390–414.
Hartman, Robert (2017). *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness*. Routledge.
Levy, Neil (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford University Press.
Lovett, Adam and Stefan Riedener (2024a). The Good Life as the Life in Touch With the Good. *Philosophical Studies*, 181(5), 1141–1165.
——— (2024b). Commonsense Morality and Contact with Value. *Philosophy and Phenomenological Research*, 109(1), 410–430.
——— (2025). Touching the Good: Special Relationships as Contact with Value. *Journal of Ethics and Social Philosophy*, 30(2), 179–204.
Mantel, Susanne (2017). Three Cheers for Dispositions: A Dispositional Approach to Acting for a Normative Reason. *Erkenntnis*, 82(3), 561–82.
Kane, Robert (1996). *The Significance of Free Will*. Oxford University Press.
Khoury, Andrew C. (2018). The Objects of Moral Responsibility. *Philosophical Studies*, 175(6), 1357–1381.

MacCormick, Donald N. (1978). The Obligation of Reparation. *Proceedings of the Aristotelian Society*, 78, 175–193.
Menges, Leonhard (2020). Responsibility and Appropriate Blame: The No Difference View. *European Journal of Philosophy*, 29(2), 393–409.
Montminy, Martin (2019). Derivative Culpability. *Canadian Journal of Philosophy*, 49(5), 689–709.
——— (2022). It Was Not Supposed to Happen Like That: Blameworthiness, Causal Deviance and Luck. *Philosophical Studies*, 180(2), 439–449.
Moore, Michael S. (1997). *Placing Blame: A Theory of the Criminal Law*. Oxford University Press.
Nagel, Thomas (1979). Moral Luck. In *Mortal Questions* (24–38). Cambridge University Press.
Nelkin, Dana K. (2011). *Making Sense of Freedom and Responsibility*. Oxford University Press.
Pereboom, Derk (2001). *Living Without Free Will*. Cambridge University Press.
——— (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press.
Reis-Dennis, Samuel (2018). Responsibility and the Shallow Self. *Philosophical Studies*, 175(2), 483–501.
Riedener, Stefan (2020). Don't Make a Fetish of Faults: A Vindication of Moral Luck. *Philosophical Studies*, 178(3), 693–711.
Robichaud, Philip and Jan Wieland (Eds.) (2017). *Responsibility: The Epistemic Condition*. Oxford University Press.
Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.
——— (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Belknap Press of Harvard University Press.
Schlick, Moritz (1939). When Is a Man Responsible? In *Problems of Ethics* (141–158). Prentice-Hall.
Schroeder, Mark (2007). *Slaves of the Passions*. Oxford University Press.
Smith, Angela M. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics*, 115(2), 236–271.
——— (2008). Control, Responsibility, and Moral Assessment. *Philosophical Studies*, 138(3), 367–392.
Sripada, Chandra (2016). Self-Expression: A Deep Self Theory of Moral Responsibility. *Philosophical Studies*, 173(5), 1203–1232.
Strawson, Galen (1986). *Freedom and Belief*. Oxford University Press.

——— (1994). The Impossibility of Moral Responsibility. *Philosophical Studies*, 75(1–2), 5–24.
Strawson, Peter (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48, 187–211.
Talbert, Matthew (2025). Blameworthiness and Causal Outcomes. *Erkenntnis*, 90, 2481–2501.
Turri, John (2011). Manifest Failure: The Gettier Problem Solved. *Philosophers' Imprint*, 11(8), 1–11.
Van Inwagen, Peter (1983). *An Essay on Free Will*. Clarendon Press.
Wallace, Robert H. (2023). Agency: Let's Mind What's Fundamental. *Philosophical Issues*, 33(1), 285–298.
Wallace, R. Jay. (1996). *Responsibility and the Moral Sentiments*. Harvard University Press.
Waller, Bruce (2011). *Against Moral Responsibility*. MIT Press.
Watson, Gary (1975). Free Agency. *The Journal of Philosophy*, 72, 205–220.
——— (1996). Two Faces of Responsibility. *Philosophical Topics*, 24(2), 227–248.
Wolf, Susan (1993). *Freedom Within Reason*. Oxford University Press.
——— (2003). Sanity and the Metaphysics of Responsibility. In Gary Watson (Ed.), *Free Will*, 2nd ed. (372–387). Oxford University Press.
Zimmerman, Michael (1986). Negligence and Moral Responsibility. *Noûs*, 20(2), 199–218.