

Towards Agent Modalist Reasons-sensitivity

Chris Cho

This is a preprint of an article forthcoming in The Philosophical Quarterly. Please cite the final published version when it appears

Abstract

Many find it plausible that acting freely requires a kind of sensitivity to reasons. Modalists define this sensitivity in terms of what S does in the relevant alternative possibilities. Agent modalists take S to be an agent, assessing their reasons-sensitivity by considering what the agent would have done in those possibilities. While intuitive, agent modalists struggle with Frankfurt cases. This has led some to a mechanism-based approach. Here, I raise two objections to this approach and defend agent modalism. I argue that (1) the mechanism approach permits composite mechanisms, leading to an unattractive disjunction, and (2) excluding the agent undermines the original appeal of reasons-sensitivity. I defend agent modalism by arguing it can handle Frankfurt cases if it holds fixed the nonoccurrence of agent-altering events. This requirement also helps address rational blind spots. Thus, I aim to rehabilitate agent modalism as a viable account of reasons-sensitivity in free action.

Keywords: reasons-sensitivity, compatibilism, modality, free will, modalism, Frankfurt cases, rational blind spots

1 Introduction

According to one popular view of freedom in the sense required for moral responsibility, acting freely requires being sensitive to reasons. What does being sensitive to reasons mean? According to *modalists*, or *modalism*, *S* being sensitive to reasons with respect to some action is roughly a matter of what *S* does in the relevant alternative possibilities. But in spelling out what ‘*S*’ refers to, modalists are divided into two camps. *Agent* modalists view *S* as an agent, so understanding an agent’s moral responsibility involves investigating what the agent would have done in the relevant alternative possibilities.¹ Agent modalism is the natural view, but many have pointed out that agent modalists cannot make sense of an agent’s reasons-sensitivity in Frankfurt cases, where an intervener prevents the agent from acting in ways that would count her as being reasons-sensitive. Some modalists have thus endorsed a *mechanism* approach to reasons-sensitivity.² On this view, an agent’s moral responsibility is a function of the modal properties of the agent’s *mechanism* that issued her action. Within the literature on reasons-sensitivity, the mechanism view is often regarded—either implicitly or explicitly—as an improvement over agent modalism, precisely because it can accommodate the challenges posed by Frankfurt cases.³

¹ Agent modalists include Fara (2008), McKenna (2013), Smith (1997, 2003), Vihvelin (2004, 2013) and others.

² Mechanism modalists include Fischer and Ravizza (1998) and Haji (1998).

³ See, for example, McKenna (2001: 93–94), Stout (2016: 402), Kaiserman (2021: 691), and Loewenstein (2023). Loewenstein appears to take for granted that the mechanism

Now, I have modalist intuitions about reasons-sensitivity, so I think *some* form of modalism is correct.⁴ However, I will argue that we should dampen our hopes for the mechanism approach. I develop two main objections. First, I argue that the mechanism view permits composite mechanisms, which leads to an unattractive disjunction: either the view collapses into agent modalism, or the mechanisms become too rigid to account for reasons-sensitivity. Second, I argue that removing agents from the picture undermines the very appeal of reasons-sensitivity.

In light of these problems, I propose that agent modalists can respond to Frankfurt cases if they evaluate the relevant counterfactuals while holding fixed the nonoccurrence of what I call *agent-altering* events. On this basis, I aim to rehabilitate agent modalism as a viable and attractive account of reasons-sensitivity.

Here is the structure of the paper. Section 2 lays out the basics of modal reasons-sensitivity. Section 3 explains how Frankfurt cases have motivated the shift from agents to mechanisms, and presents two objections to the mechanism view. Section 4 shows how agent modalism can accommodate Frankfurt cases by appealing to agent-altering events, and argues that this same strategy addresses the challenge of rational blind spots.

approach is superior.

⁴ For non-modalist approaches to reasons-sensitivity, see Sartorio (2016) and Heering (2022a).

2 Modal Reasons-sensitivity

Consider an addict and a non-addict who are each presented with a drug in front of them. Both decide to take it. It seems right to judge that while the non-addict took the drug freely, the addict did not. How can this difference in freedom be explained?

Modalists offer an intuitive explanation. Here is the basic template. Suppose the non-addict learns that the drug would cause the worst headache of his life. It's plausible he would refrain, whereas the addict, because of his addiction, would still take it. In other words, given sufficient reasons to refrain from taking the drug, while the non-addict would have responded to those reasons and acted accordingly, the addict would have failed to respond in the same way. Thus, there is a difference in their *modal profiles* with respect to their drug taking, which grounds the differences in their freedom.

Given this basic template, modalism varies across two axes. The first concerns the kind of modal profile a reasons-sensitive agent must have. For Fischer and Ravizza (1998), a modal profile suffices for reasons-sensitivity provided there is at least one world in which the agent does otherwise given a sufficient reason to refrain. But many have pointed out that this condition is too weak and would make too many agents reasons-sensitive.⁵ Surely, even our severe addict would not take the drug if he knew that

⁵ See, for example, McKenna (2005) and Todd and Tognazzini (2008).

doing so would somehow start a zombie apocalypse!

Perhaps the right kind of modal profile for reasons-sensitive agents, then, is one where they refrain from acting in *every* world in which they are given sufficient reason to refrain. But this ‘perfectionist’ modal profile would preclude almost all agents from being morally responsible. For instance, consider a weak-willed agent who steals from the supermarket even though she knows that she shouldn’t do so. She is morally responsible for stealing, but the perfectionist model does not deliver that verdict. Even when she recognizes reasons to refrain, she does not always do so.

A nice balance between the two types of modal profiles has been to say that reasons-sensitive agents don’t act in a *sufficient proportion* of worlds in which sufficient reasons to refrain are present.⁶ In my view, the most plausible form of reasons-sensitivity ought to endorse this clause. Of course, it’s difficult to spell out what exactly would amount to a *sufficient* proportion of worlds. All I will say for this paper is that there is an intuitive sense in which the non-addict passes a certain threshold for sufficiency while the addict does not.⁷

Now, another way in which modalism can vary is in the *type* of reasons we test our agent against. So far, we’ve only tested our agent’s behavior in alternative possibilities given one type of reason—sufficient reasons to *refrain*. In other words, we have only looked at worlds in which the agent

⁶ This is essentially what Brink and Nelkin (2013) call ‘Moderate Responsiveness’.

⁷ See Manley and Wasserman (2008) for an attempt on this issue. Note that they say ‘suitable’ instead of ‘sufficient’ proportion. This is merely a terminological difference.

has sufficient reason to do otherwise. But this raises the question: How do these agential modal properties tell us whether the agent was sensitive to the reason in the *actual* world? Given that the agent ϕ -ed in the actual world, the modalist we have on the table so far only tells us what the agent would have done in light of reasons to *not*- ϕ . How can the set of such modal properties tell us that the agent was sensitive to reasons *to* ϕ in the actual world?

Modalists might ask us to consider McKenna's (2013) thermostat example. If we want to know whether a thermostat is well-functioning or not, it is not enough to look at the thermostat and realize that the temperature it is pointing to matches the room temperature; after all, a broken clock is still correct twice a day. Instead, the natural way for us to test whether it is well-functioning or not is to see how it would behave if it were set to a different temperature. For example, if the room temperature is currently at 20° and I were to adjust the thermostat to 30°, would the room temperature change to 30°? If the answer is yes, we would call that thermostat well-functioning. In other words, a thermostat is deemed well-functioning on account of certain modal properties about how it would have behaved given different temperature inputs. Similarly, the modalist may argue, figuring out how the agent would behave in the face of reasons different from the reason in the actual world is informative in telling us whether the agent was sensitive to the actual reason.

One might still worry here. What if a reasons-sensitive agent in many

of the alternative possibilities acted in a way that the reason recommended but didn't do it *for that reason*? Let me explain. Suppose that the following counterfactual is true:

(A) If the non-addict realized that the drug was headache-inducing, the non-addict would not take the drug.

So, in a sufficient proportion of worlds in which the non-addict realized that a drug was headache-inducing, the non-addict did not take the drug. But now suppose that in most of those worlds in which the non-addict realizes that the drug is headache-inducing, he also sees that his stove is on fire. Surprised, our non-addict runs to put the fire out. In this case, what makes him refrain from taking the drug in those worlds is presumably not because the drug is headache-inducing, but because the stove is on the fire. This leads to an interesting result. The counterfactual is true, but the reason why the non-addict does not take the drug is not because of the antecedent, but for some other reason.⁸

This is problematic for those modalists who analyze reasons-sensitivity by looking at only those worlds in which the agent has reasons to refrain. This, then, motivates an additional requirement on reasons-sensitivity. Instead of looking at those worlds in which the agent has sufficient reasons to refrain, modalists can expand their search to worlds in which the agent has sufficient reason to *do* the action they did in the actual world. More specifically, they can hold fixed the actual reason that led an agent to ϕ and

⁸ For worries of this kind, see Loewenstein (2023: 132–133) and Fischer and Ravizza (1998: 64–65).

vary *everything else*.⁹ This would mean that the agent has, in addition to the modal properties we've discussed so far, a different set of modal properties. How would these modal properties help us in understanding whether the agent was sensitive to the actual reason? Consider our thermostat again. Here is a different way to test whether it is well-functioning. Instead of seeing whether the room temperature covaries with what the thermostat indicates, we set the thermostat to maintain a particular temperature and see whether the room temperature remains constant when we vary all the other variables that might have been causing the room to be at a certain temperature. For example, perhaps the room was designed to be at different temperatures based on how many people are in the room—say, 20° when there is just one person, 25° when there are two, and so on. As it happened, the thermostat was pointing to 20° when I was alone in the room. In that case, if we were to vary the number of people in the room, we could tell if the thermostat was a well-functioning one. Likewise, then, for our agent. By seeing whether the same reason would have produced the same outcome in the relevant worlds, we can see whether the agent was sensitive to the actual reason.

In sum: Modalism varies across two different axes: in terms of what the right kind of modal profiles are, and in terms of what kind of reasons we test our agent against. To be clear, these are not mutually exclusive options. In fact, while most modalists have only incorporated the first, I think that

⁹ Of course, I don't mean *everything*. We have to hold fixed certain background conditions.

the most plausible kind of modalism (both agent- and mechanism-based) is one that incorporates both. That is, a view that looks at worlds in which the agent has sufficient reason to refrain as well as the same reason as they had in the actual world. In light of this, then, I take the following to be the most plausible modalist view of reasons-sensitivity:

RS: an agent is reasons-sensitive with respect to ϕ -ing iff the agent ϕ -s according to sufficient reasons for and against ϕ -ing in a sufficient proportion of worlds where such reasons are present.

Obviously, this is an agent-based approach to reasons-sensitivity. The object of reasons-sensitivity is an *agent*. We hold fixed the agent across different worlds. But agent-based approaches face a worry. Those who are sympathetic to spelling out freedom in terms of reasons-sensitivity are often compatibilists who find Frankfurt cases convincing. And one thing that is clear in Frankfurt cases is that agents in such cases acted freely, which would mean that they were reasons-sensitive. But under **RS**—indeed, on any agent-based approach—such agents don’t turn out to be reasons-sensitive. In the next section, I spell out this worry further, motivate the mechanism approach, and put forth two worries for mechanism modalists.

3 Frankfurt Cases and the Mechanism Approach

There are many different versions of Frankfurt cases. Here is one I like:

Frank and Furt: a neuroscientist has been secretly monitoring the brain processes of an agent, call him Frank, who is deliberating

whether to shoot Furt or not. The neuroscientist can reliably predict the choices that Frank is about to make by looking at the activity in his brain, and can also manipulate Frank's brain in a way that guarantees that Frank will decide to shoot Furt. He plans to intervene if he predicts that Frank will not shoot Furt on his own. As it happens, Frank does decide to shoot Furt, motivated by his own reasons, and without the intervention of the neuroscientist. (Sartorio, 2016)

These cases supposedly falsify the Principle of Alternative Possibilities (PAP), according to which moral responsibility requires one's ability to have done otherwise (Frankfurt, 1969). Admittedly, it's a matter of controversy whether Frankfurt cases do falsify PAP.¹⁰ But one thing that everyone on both sides of the debate seems to agree on is that Frank is morally responsible for his decision to shoot Furt. That much, is uncontroversial.

But if modalists think moral responsibility requires reasons-sensitivity and endorse **RS**, they cannot accommodate this uncontroversial claim. Suppose that one sufficient reason for Frank not to shoot Furt is his realization that Furt is actually his long-lost brother. So, one of the relevant counterfactuals in analyzing Frank's reasons-sensitivity is this:

(F): If Frank had realized that Furt was his long-lost brother, Frank would not have decided to shoot Furt.

Under standard counterfactual semantics, (F) is true just in case the consequent is true in worlds most similar or closest to the actual world in which the antecedent is true.¹¹ The problem is that in most of the

¹⁰For an overview of the debate, see Sartorio (2017a).

¹¹I am, of course, relying on Lewis's (1973) approach.

similar worlds in which Frank realizes that Furt is his long-lost brother, the neuroscientist will be present in those worlds and would make Frank decide to shoot Furt anyway despite the reason to refrain. After all, the most similar worlds are those in which the antecedent is true with *minimal changes* from the actual world. Thus, (F) is false. But by **RS**, (F) should be true if modalists wish to maintain that Frank is morally responsible.

Given this problem, agent modalists face a dilemma: either they have to somehow show that Frank is in fact reasons-sensitive or admit that reasons-sensitivity is not required for moral responsibility. In Section 4, I will argue that the first option is viable. For now, I shall discuss how some have endorsed a third option: a *mechanism* approach to reasons-sensitivity.

On this approach, the object of reasons-sensitivity shifts from the agent to the agent's *mechanism*, which is nothing over and above the 'process that leads to the relevant upshot.' (Fischer & Ravizza, 1998: 38). To understand whether an agent is morally responsible or not, then, instead of looking at how the *agent* would have behaved in alternative possibilities, we see how their *mechanism* would have behaved in alternative possibilities. That is, we hold fixed the mechanism that was operative in the actual world (the actual operative mechanism) and see how it would have behaved in light of different (or the same) reasons.

One primary advantage of this approach—indeed, the primary motivation for Fischer and Ravizza, who first proposed this—is that it allows modalists to make agents in Frankfurt cases reasons-sensitive. Here is

how. One difference between Frank's decision-making process in the actual world and one in which the neuroscientist intervenes is that in the latter, the neuroscientist is part of that causal process. So, if a mechanism refers to the process that leads to the action, it's clear that the actual operative mechanism is different from the one that is operative in worlds in which the neuroscientist intervenes. And in trying to find what the relevant worlds are, mechanism modalists hold fixed the actual operative mechanism, which does not involve the neuroscientist (or any other intervener). Hence, worlds in which the neuroscientist is present are screened off from the set of relevant worlds. The result is that while the agent is not reasons-sensitive, the relevant mechanism that issued their action is reasons-sensitive. And moral responsibility requires that the mechanism be reasons-sensitive, not the agent.

3.1 Problem of Mechanism individuation

Despite its ability to handle Frankfurt cases, the mechanism approach faces two worries.¹² First is the problem of mechanism individuation. One crucial feature of this approach is that we hold fixed the actual operative mechanism. But this means that we should have some principled basis for what counts as the same mechanism across worlds. Unfortunately, Fischer and Ravizza concede they do not have a strict criterion. They claim, 'It is

¹²For simplicity's sake, I will proceed with Fischer and Ravizza's (1998) approach in mind, though my worries carry on to Haji's (1998) approach too.

simply a presupposition of this theory as presented here that for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in action...' (47). Critics have pressed them on this point:

I find the phrase "kind of mechanism that issues in action" of little help by itself in suggesting the features that are relevant. More helpful would be a taxonomy, even a rough one, of the things that Fischer and Ravizza regard as the relevant "kinds of mechanisms," but none is provided. They do suggest that an action-producing mechanism that is not responsive to reasons at all must be different in kind from one that is responsive to reasons, and one that is responsive to moral as well as prudential reasons must be different in kind from one that is responsive only to prudential reasons. But apparently this sort of consideration does not individuate kinds of mechanisms as finely as Fischer and Ravizza wish to do. (Ginet, 2006: 234)

It is an indispensable part of their theory that we are to test how the same mechanism behaves in response to different reasons. How are we to settle whether, when different reasons are put to an agent, the same mechanism is operative? We need some purchase on what it is that we are holding fixed when we hold fixed "the same mechanism" while testing its degree of responsiveness. (McKenna, 2013: 161)

Admittedly, a theory having a vague concept is no reason to reject the theory. Many foundational concepts in philosophy and science are vague. For example, in discussing personal identity, the notion of personhood is often left under specified; yet no one dismisses a view of personal identity for *that* reason. Similarly, biologists often define the concept of a 'species' as a group of organisms that can interbreed and produce fertile offspring. But

this does not accommodate cases of hybrid species, asexual reproduction, or ring species. And no one rejects biology because it doesn't!

Nevertheless, in my view, the particular vagueness surrounding Fischer and Ravizza's notion of a mechanism leads to troubling results that warrant a rejection of their view. Consider the following two cases:

Case 1: Sno decides to help a stranger who has dropped their groceries in the middle of a busy street. She notices the spilled groceries and pauses to assess the situation. She considers the stranger's apparent difficulty in retrieving them, evaluates her own schedule to confirm she has enough time to assist, and reflects on the moral significance of helping someone in need. After weighing these factors, she concludes that helping is the right course of action and decides to act.

Case 2: Sno decides to help a stranger who has dropped their groceries in the middle of a busy street. Upon seeing the situation, Sno feels a sudden pang of empathy and an overwhelming sense of urgency to help. She doesn't stop to consciously deliberate about her schedule or the moral implications of her actions. Instead, she is moved by her immediate emotional response—the stranger's distress tugs at her heart, and she instinctively rushes to assist.

In both cases, Sno decides to help a stranger, but the processes—that is, the mechanisms—leading to that decision differ. The mechanism in Case 1 seems to be a *deliberative reasoning mechanism*, since Sno is reasoning through whether she can and wants to help the stranger. It also seems right to say that this mechanism is reasons-sensitive. If she had sufficient reasons not to help the stranger—say, she realized the stranger dropped an apple, to which she is extremely allergic—the mechanism would output a

decision not to help. In contrast, the mechanism in Case 2 does not seem to be reasons-sensitive. Since Sno's decision in Case 2 is a result of her emotions, let us call this the *affective mechanism*. It seems likely that in most of the worlds where an apple is present this particular mechanism outputs a decision to help.

Consider now Case 3:

Case 3: Sno decides to help a stranger who has dropped their groceries in the middle of a busy street. Upon seeing the situation, she immediately feels a strong pang of empathy, which sharpens her attention to the situation. Her emotional response highlights the stranger's visible frustration and the chaotic scene, making it clear that assistance is needed. At the same time, she considers her own schedule and whether she has enough time to help. Her empathy deepens her understanding of the stranger's difficulty, making the moral importance of helping more salient. In turn, her reasoning structures her emotional response, helping her evaluate the urgency and practicality of offering assistance. Together, her emotional and rational processes converge, leading her to conclude that helping is both the right and feasible thing to do. Feeling both convinced and emotionally moved, Sno decides to help.

In Case 3, both Sno's affective response and rational deliberation work in tandem. But this is a problem. We now have two mechanisms that together constitute the process that leads to the relevant upshot. But we have seen that while the deliberative mechanism is reasons-sensitive, the affective mechanism is not. But if both are involved in the process, how can we pick out what the relevant mechanism is so that we can test Sno's reasons-sensitivity?

One natural response is to argue that the relevant mechanism is the composite mechanism constituted by these two mechanisms, call it, the *affective reasoning mechanism*, where its reasons-sensitivity depends on the extent to which the affective mechanism impinges upon the deliberative reasoning mechanism. I have in mind Romeo who, so overwhelmed by grief and despair, fails to reason accurately about Juliet's condition, and decides to take his own life.

However, this move exposes a deeper issue: understanding the composite mechanism of affective reasoning necessitates an examination of the agent's broader psychological constitution. Emotional states such as anger, fear, or empathy and how they interact with deliberation are intricately tied to the agent's overall value system, personality, and moral framework. For instance, anger in response to perceived injustice is not merely a fleeting emotional state—it reflects the agent's deeper sense of fairness, identity, and moral commitments. Consequently, the affective reasoning mechanism cannot be fully understood without accounting for the agent's psychological makeup. So, subjecting the mechanism to alternative reasons requires holding fixed not just the mechanism itself but also the relevant aspects of the agent's broader constitution. I predict this requires us to hold fixed most, if not all, of the agent. But now, the mechanism view becomes practically indistinguishable from an agent-based view. After all, when agent modalists hold fixed the agent, they are holding fixed how the agent *is*, not necessarily their physical makeup. So, it seems, the mechanism view

collapses into an agent-based view.¹³

Defenders of Fischer and Ravizza may object. Recall that a mechanism is ‘a process that leads to the relevant upshot’ and as such, they may say, it refers to the particular mental states that led to the outcome in the actual world. So what we hold fixed is, to use a typographical distinction, [empathy], the agent of which is Sno and the object the stranger (I’m also assuming [empathy] to be temporally indexed). What we do not hold fixed, say these defenders, is empathy as a general mental state. For how a general mental state operates depends on how the agent is, leading to the collapse. And if we hold fixed [empathy] instead of empathy across possible worlds, surely we do not need to bring the agent with us. [Empathy] and other [] states can be held fixed independently of how the agent operates.

This response risks overindividuating mechanisms to the point of undermining their coherence. If the mechanism is narrowly defined as the particular mental states present in the actual world, it becomes rigid and inflexible. For example, if an agent acted out of anger in the actual world, the mechanism would likely produce the same outcome in worlds where there are sufficient reasons to act otherwise, since, after all, it is the series of [] states that led to the actual world outcome. To be sure, mechanisms can

¹³Does it collapse only in cases where we have a composite mechanism? I don’t think so. Presumably, counterfactual claims about an agent’s emotional response to a stimulus are also intricately tied to the agent’s psychological makeup. So, even in cases where we do not seem to have competing mechanisms, it is possible to argue that the talk of mechanisms requires talking about the agent as a whole. I do, however, think that the collapse is more evident in composite mechanism cases.

be inflexible; those that aren't reasons-sensitive are said to be so. But their inflexibility should stem from how the mechanism responds to reasons given a circumstance, not from how they are constituted.¹⁴

In sum: The inherent vagueness in Fischer and Ravizza's notion of mechanism individuation allows cases of there being composite mechanisms, the constituents of which are inconsistent in their reasons-sensitivity. This is not problematic for Fischer and Ravizza, since they may simply construe the composite mechanism as a single mechanism. But the consequence of this response is an unattractive disjunction: Either the mechanism view collapses into agent modalism or their mechanisms lose the flexibility required for reasons-sensitivity. This problem may be resolved if there is a way to individuate mechanisms such that they aren't too coarse to collapse into agent modalism and too fine to be inflexible. But I predict that to be a difficult task.

¹⁴One might wonder whether Fischer and Ravizza will even accept the defense in the previous paragraph as a defense, given that they argue that mechanisms ought to be 'temporally intrinsic' (46), meaning that they should not entail the action they are supposed to cause. Perhaps so. But this requirement can easily be satisfied. One may construe the mechanism in Sno's case as one that involves all the [] states *except* the very last [] state that causes Sno's decision. Now, assume that causation is intransitive (which I don't take to be controversial. See, for example, Hitchcock (2001), Yablo (2002), Sartorio (2006), Hall (2007)). Surely, the penultimate [] state causes the excepted [] state, which in turn causes Sno's decision. On this construal, the mechanism satisfies the intrinsic requirement.

3.2 *The switch from an agent to a mechanism*

Here is another worry for the mechanism approach: It drops the agent out of the picture. On the standard picture of reasons-sensitivity, an agent is morally responsible in virtue of them being reasons-sensitive. But in the mechanism approach, an agent is morally responsible in virtue of their mechanism being reasons-sensitive. There is something very odd about this claim and this move is at least *prima facie* dubious. Consider how we often excuse others. We often assert claims like ‘Don’t blame her, she couldn’t help it’. The incapacity to be sensitive to reasons that we are attributing here is to the agent, not to her mechanism. If we adopt the mechanism approach, our everyday practices would simply be a useful paraphrase for mechanism talk. That may indeed turn out to be true, of course. But a theory that does not require a conceptual re-thinking of our everyday discourse would be theoretically advantageous than the one that does (Nelkin, 2011: 19).

Relatedly, it’s unclear whether the switch from an agent to a mechanism preserves the initial motivations for having reasons-sensitivity as a requirement for moral responsibility in the first place. Recall *how* the addict case motivated the idea of moral responsibility requiring agents to be reasons-sensitive. There, whether the non-addict or the addict was sensitive to reasons or not tracked our judgments of moral responsibility. That then provided grounds for the idea that one must be sensitive to

their reasons to be morally responsible. If the agent was not responding to a particular reason, we intuited that the agent is a particular kind of an insensitive agent such that we excuse their moral responsibility. In other words, the idea of moral responsibility requiring reasons-sensitivity gains initial momentum because we find morally responsible agents to be *themselves* reasons-sensitive, whereas we deem non-responsible agents to be insensitive to reasons. But if we switch to talking about mechanisms instead of agents, the intuitive connection between whether an agent is morally responsible and whether *they* are reasons-sensitive is no longer present. The point is that dropping the agent out of the picture divorces the mechanism view from modalism's initial attractions. It may be wise, then, to have a modalist view where we keep the agent in place.

To be fair, Fischer and Ravizza (1998) do not drop the agent *completely*. To establish a relation between a reasons-sensitive mechanism and a morally responsible agent, they argue that agents must *own* their reasons-sensitive mechanism, for which they offer three necessary and sufficient conditions: (1) an agent must see themselves as the source of his behavior, (2) they must accept that they are a fair target of the reactive attitudes as a result of how they exercise their agency, and (3) their view of themselves is based in an appropriate way on the evidence (214). But, in addition to the problem of mechanism individuation, many have taken this ownership condition—which trivially is not required for agent modalism—to be problematic. For example, both McKenna (2000) and Ginet

(2006) note that (2) leads to an extreme subjectivist position, where a hard determinist who believes that determinism is true and incompatible with responsibility will be left off the hook because they don't satisfy (2). Surely whether an agent is responsible or not should not depend on how they think of themselves!

There are other objections to the ownership condition, but I will not discuss them here.¹⁵ I simply note them to make the point that this ownership condition isn't something that is required for the agent modalist. While a mechanism modalist may try to offer plausible responses to the objections and even succeed in doing so, it seems to be theoretically advantageous to uphold a view that does not have to spend time dealing with such responses in the first place.

In sum: Modalists should be wary of endorsing the mechanism approach. Note that the problems I've raised in this section all surround the notion of a mechanism. If we set aside issues about Frankfurt cases for agent modalism, then agent modalism has the upper hand. It's relatively uncontroversial what makes an agent the same agent across worlds, and we retain the initial motivation for reasons-sensitivity being necessary for moral responsibility. Perhaps mechanism modalists have a response to the problems I've raised, but even if so, it would be better if we uphold a view that is less riddled with roadblocks.

¹⁵See, for example, Yaffe (2000: 432–433), Watson (2001: 386–392), Russell (2002: 595–599)

4 Towards Agent Modalism

Let us now turn back to agent modalism. Previously, I mentioned that the agent modalists face a dilemma in the face of Frankfurt cases: either they have to somehow show that Frank is in fact reasons-sensitive or admit that reasons-sensitivity is not required for moral responsibility. In this section, I shall take up the former task. Let us begin with two stories.

On August 1, 1966, Charles Whitman entered the University of Texas at Austin and carried out a mass shooting, having already killed his mother and wife earlier that day. He killed a total of 17 people, leaving 31 wounded. Whitman was not a born killer. Before the incident, he noted in his diary an overwhelming desire to kill, but expressed confusion about the source of these urges. He even requested an autopsy if he were killed during the event. Alas, the autopsy revealed a pecan-sized tumor pressing against Whitman's amygdala, a brain structure crucial for regulating fear and aggression (Wallenfeldt, 2016).

Similarly, consider the famous case of Phineas Gage. On September 13, 1848, while working on a railroad, a large iron rod was accidentally driven through Gage's left frontal lobe, a brain structure known to regulate personality and behavior. Surprisingly, Gage survived. Those around him, however, noticed a drastic change in his personality. Before the accident, he was described as hardworking, responsible, and a well-balanced individual. Afterwards, he became irreverent, fitful, and unable to adhere to social

norms and expectations. As his acquaintances collectively proclaimed: He was no longer Gage (Gearhart, 2013).

These events illuminate an interesting fact: Certain events alter who we are as individuals. So let's say that both Whitman and Gage experienced what we might call *agent-altering* events—incidents that fundamentally change what one values and how one exercises their mental faculties. That agent-altering events occur, I presume, is not controversial.

With this in mind, recall why Frankfurt cases troubled the agent approach. Here is (F) again:

(F): If Frank had realized that Furt was his long-lost brother, Frank would not have decided to shoot Furt.

To analyze this sentence, agent modalists had to search through nearby worlds in which the antecedent is true while holding the relevant facts about the agent fixed. But the problem was that the criteria for what makes a world *nearby* was based on comparative similarity with the actual world, in which case the neuroscientist will be present in the nearby worlds.

To remedy this problem, one may replace the notion of comparative similarity in favor of another account to handle what makes a world nearby or relevant. But I find the idea of comparative similarity rather intuitive. As such, I propose that agent modalists ought to fix further facts other than the facts about the agent. Specifically, my view is that, at least when it comes to assessing counterfactuals relevant to reasons-sensitivity, we should fix the nonoccurrence of agent-altering events when searching through the possible worlds.

For this requirement to allow agent modalists to make sense of (F), the neuroscientist's intervention must be considered an agent-altering event. It's quite plausible to see that it is. Consider Frank* who is just like Frank except he realizes that Furt is his long-lost brother. Noticing this, the neuroscientist hiding in the bushes presses his button, changing Frank*'s brain chemistry so that his ability to recognize and follow through with his reasoning is tampered with, similar to the effect the tumor had on Whitman. Now, suppose we took a snapshot of Frank*—roughly, the set of facts pertaining to his psychology—at the point of this alteration and a snapshot of ordinary Frank. There would be a significant difference between them with respect to their fundamental reasoning capabilities. In particular, Frank* *qua* his snapshot is an agent such that even in a world where he realizes Furt's identity and *there is no neuroscientist*, he decides to shoot Furt. The point is that just like how the tumor is an agent-altering event for Whitman, the neuroscientist's intervention also acts as an agent-altering event for Frank. By if so, when evaluating (F) by considering nearby possible worlds, those in which the neuroscientist intervenes would be excluded, and (F) would be deemed true—*mutatis mutandis* for other counterfactuals of its kind.

I anticipate some objections. Here is the first one: If the intervention causes an agent-altering event, do agent modalists even need the requirement that we have to fix the nonoccurrence of such events? If Frank and Frank* are just different agents, in the relevant sense, don't intervention

worlds get screened out because nearby worlds ought to include the same agent?

I have two things to say in response. First of all, this objection assumes that altering an agent creates a different agent altogether. That's not necessarily the case. While Gage underwent profound changes, Whitman's acquaintances still recognized aspects of Whitman's personality such that he was still, in some sense, *Whitman*. Second, even if agent-altering events were such drastic events, fixing the agent in conjunction with comparative similarity will not guarantee that nearby worlds are not intervention worlds. To see why, consider a world in which the antecedent of (F) is true. The time frame for this world is as follows: At time t_0 , Frank realizes that Furt is his brother. At t_1 , the neuroscientist presses a button, altering Frank's brain chemistry. Now, if we were to take a snapshot of Frank at t_1 , we would have an altered agent from Frank in the actual world, since his faculties have now been compromised and altered. In contrast, if we were to take a snapshot of Frank at t_0 , the relevant facts of *that* snapshot will not be different from that of Frank in the actual world. The objection only goes through if the set of agential facts true in the actual world is compared with the set of agential facts captured at t_1 . But there is no reason to think why it should be. The set of facts can just as much be compared to the set of facts captured at t_0 . In fact, I am inclined to think that this is the correct comparison in the first place. After all, what occurs at t_0 is precisely what is dictated by the antecedent of (F). So, even though it is true that *if* the

neuroscientist were to intervene, we would have an agent-altering event, that does not make the requirement I am proposing here superfluous.

Here is the second objection. One might argue that my appeal to agent-altering events suffers from a similar problem that I raised for the mechanism approach. While I argued that the mechanism view faces serious difficulties demarcating which processes count as “the same mechanism” across worlds, one might think the same issue arises for individuating agent-altering events. After all, though cases like Phineas Gage and Charles Whitman are paradigmatic cases of agent-alteration, many interventions or shifts in mental states—such as slight chemical tweaks, changes in motivational salience, or minor psychological pressures—will fall into a gray area. Without a principled way of drawing the line between agent-altering and non-agent-altering events, the worry goes, my proposal would face the same kind of vagueness as the mechanism approach.¹⁶

Again, I have two responses. First, I concede that the agent-based approach, like the mechanism-based one, faces a line-drawing problem—there will be clear cases of agent-altering events, and others where it’s unclear. But this parallel should not be overstated. My objection to the mechanism-based approach is not simply that it involves vagueness. Rather, it’s that the vagueness surrounding how to individuate mechanisms, together with the possibility of composite mechanisms, produces an unstable framework. Specifically, as discussed in Section 3.1, it leads to an unattractive disjunc-

¹⁶I thank both of my anonymous reviewers for raising this objection.

tion: either the view collapses into agent modalism (by letting mechanisms track the whole agent), or the mechanisms become too rigid to support reasons-sensitivity. The agent-based approach does not face this. On this view, what amounts to a composite mechanism just *is* the agent. So, while there is a line-drawing problem, this does not itself threaten the coherence or integrity of the view in the way it does for the mechanism-based account.

Second, the mention of slight or minor interventions does raise a potential challenge for my view—namely, the possibility of a Frankfurt case in which the intervention is not an agent-altering event because it is so minimal that it fails to fundamentally change what one values or how one exercises one’s mental faculties. Could such cases exist? Perhaps. But I am doubtful that they pose a real problem for the reasons-sensitivity condition. Recall that in a Frankfurt case, given that the agent had a reason to ϕ in the actual world, the intervener steps in when there is a reason to not- ϕ . Importantly, we see with **RS**, this reason must be a *sufficient* reason to not- ϕ . Now, I suspect that as we weaken the strength of the intervention, we correspondingly weaken the sufficiency of that reason. Thus, when the intervention is so minimal, the intervention worlds get screened out from the reasons-sensitivity analysis because they aren’t worlds where the agents have a sufficient reason to not- ϕ in the first place.

Here is an example to illustrate this point. Suppose Frank is deliberating about whether to steal from a bank, and he ultimately decides to do so for self-interested reasons. Further suppose that in a nearby world,

Frank begins to reflect on the idea that life is not all about money, which constitutes a reason to refrain from stealing. However, the intervener is poised to ensure that Frank steals. So, upon detecting Frank's moment of moral hesitation, the intervener subtly increases Frank's egoism—just enough to tip the scales and keep him committed to stealing.

Is this a case where Frank fails to respond to a *sufficient* reason to refrain from stealing? I don't think so; or, at least, it's unclear. By stipulation, the intervention is—and must be—minimal enough to be unequivocally not an agent-altering event. But that requirement raises a problem: if such a slight nudge is enough to ensure that Frank still decides to steal, then it's reasonable to think that he was already the kind of agent who, in a sufficient proportion of worlds where he has a reason to refrain, would still decide to steal. And if so, then the reasons to refrain in those worlds may not be genuinely sufficient after all. In short, as we weaken the intervention to avoid it being agent-altering, we also diminish the plausibility that the agent is failing to respond to a truly sufficient reason—and thus weaken the plausibility that this case threatens the reasons-sensitivity condition on the table. Conversely, to ensure a failure to respond to a sufficient reason, the intervention would likely need to be stronger, which in turn risks counting as agent-altering.

Here is the final objection.¹⁷ I've argued that the kind of reasons-sensitivity that the agent modalist relies on to make sense of Frankfurt

¹⁷I thank an anonymous reviewer for raising this point.

cases requires holding fixed the absence of agent-altering events. But, one might say, it seems that what makes an agent's snapshot in one world significantly different from its counterpart in another world—the very basis for identifying an agent-altering event—depends on a comparison of modal profiles. So, am I not appealing to a reasons-sensitivity analysis in order to identify when reasons-sensitivity should be applied? Doesn't that amount to circularity?

My response is: no. The agent-altering part of my view does not involve circularity because it bottoms out in a simpler kind of counterfactual analysis—one that is prior to the more demanding one used to assess moral responsibility. In particular, the kind of analysis needed to evaluate an agent's reason-sensitivity requires that we exclude agent-altering events; but what counts as an agent-altering event bottoms out in a simpler counterfactual analysis that is different in kind than the one utilized to analyse reasons-sensitivity.

Here's an example to illustrate why this is the case. Consider Frank and Frank* again and their 'snapshots'. Their modal profiles differ. While Frank ϕ -s in response to both a reason to ϕ and a reason to not- ϕ (because the counterfactual intervener ensures that ϕ occurs), Frank*, by contrast, refrains from ϕ -ing in response to the same pair of reasons. This modal difference reflects a relevant difference in structure, one that I call agent-altering. And in obtaining this difference, we do not hold fixed the absence of agent-altering events. Thus, the kind of counterfactual analysis required

for assessing an agent's reasons-sensitivity does not presuppose the one utilized for identifying agent-altering events. There is no circularity problem.

4.1 *Rational blind spots*

I believe I have given the resource to agent-based modalists to make sense of Frankfurt cases. To further motivate this requirement, I shall now discuss how it allows them to respond to another set of problematic cases, proposed by Heering (2022b). Here they are:

Jealousy: Razvan is a well-tempered, rational man. He is responsive to a range of reasons for and against staying with his spouse related to their fidelity. This involves correctly reading social cues indicative (and not indicative) of fidelity, correctly assessing evidence for and against infidelity, basing his decisions about his relationship on such evidence, and finally perhaps also weighing the severity of breaches of truth should they occur. But Razvan suffers from a blindspot. When he sees his spouse talking to other people in a certain way (smiling at them while tilting their head), Razvan is overcome with great jealousy that prevents him from seeing sufficient reasons pertaining to the fidelity of his spouse. One day, Razvan sees his spouse talking to a person they know (smiling, tilting their head). Overcome with jealousy, Razvan decides to break up with his spouse.

Trauma: Due to an old, entrenched trauma, Katharina cannot translate reasons for and against verbally abusing people into action properly. Whenever she is faced with a situation in which the relevant reasons become pertinent, she ends up verbally abusing someone. This trauma applies to everyone, except David, with whom Katharina has formed a somewhat mysterious bond. One day David is being very annoying. Katharina

assesses reasons for and against verbally abusing him, and, to her surprise manages not to let lose at David.

In Jealousy, the right thing to say is that Razvan was not being reasons-sensitive. But consider what he would have done in the nearby worlds. Certainly, in the nearby worlds Razvan's spouse need not smile and tilt their head in a particular way. Indeed, since that is a highly detailed fact about that world, it easily would not obtain in the nearby worlds. But if so, Razvan will no longer be affected by his blindspot and will be able to act according to what his reasons recommend. According to **RS**, however, that means Razvan will be considered as reasons-sensitive!

In Trauma, we have the opposite situation. It seems right to say that Katharina is sensitive to her reasons in the actual world. But in the nearby worlds, surely Katharina need not be talking to *David*. And in those worlds, she will not be able to act according to what her reasons recommend. So, by RS, Katharina will *not* be reasons-sensitive in not letting lose at David.

Heering's point is this. In these 'rational blind spot' cases, we find agents that suffer from highly specific local blockages in their psychology that prevent them from correctly reacting to a particular type of reason. But because these blockages are triggered only in a very narrow range of cases, it is thought that the agents' general modal properties are left intact, making modalism deliver the wrong verdict.

I should note that Heering targets modalists who endorse a certain 'test-cases' clause in their account of reasons-sensitivity. This additional

clause is inspired by the dispositions literature, where getting the correct analysis of a dispositional property via counterfactuals requires looking at worlds (i.e., test-case worlds) in which there is no interference to the manifestation of that disposition.¹⁸ For example, there is a sense in which a glass wrapped in bubble wrap is fragile such that if it were thrown against a wall, it would shatter. But in the nearest world in which it is thrown, the wrap prevents it from shattering. So if a disposition is based on what such counterfactuals say, we get the result that the glass is not fragile. The notion of ‘test-cases’ is meant to remedy this problem by looking at worlds in which the glass is thrown *without* the bubble wrap, since that interferes with the manifestation of the glass’s fragility disposition.

One might think that my proposal to fix the nonoccurrence of agent-altering events is the same as the test-cases clause. But that would be a mistake, and their difference bears out in how they analyze Jealousy and Trauma

In Jealousy and Trauma, invoking test-cases does not help the modalists. In Jealousy, if the smiling and tilting are the bubble wraps of Razvan’s capacity to be sensitive to reasons, they will be removed in the nearby worlds, which gets the modalists the wrong result that Razvan is reasons-sensitive. Conversely, in Trauma, it’s unclear what the bubble wrap here

¹⁸ As far as I know, Heering (2022b) is the only one who uses the ‘test-cases’ locution. Different authors offer different labels for the same general idea: Fisher’s (2013) ‘auspicious circumstances’, Mumford’s (1998) ‘ideal conditions’, Bird’s (1998) ‘normal circumstances’, Malzkorn’s (2000) ‘normal conditions’, Gundersen’s (2002) ‘standard conditions’, Cross’s (2005) ‘background conditions’, and Choi’s (2009) ‘ordinary conditions’.

is, especially since Katharina's trauma is intrinsic to her. Perhaps the interference is the presence of David, but then, like Jealousy, removing that feature in the nearby worlds gets the modalists the wrong result that Katharina is *not* reasons-sensitive.

Turn now to my proposal. Jealousy first. Consider a snapshot of Razvan in the actual world and a snapshot of Razvan in a world in which there is no smiling and tilting. Those snapshots differ since in the former, a blind spot has been activated, while in the latter, it has been deactivated. And the difference between blindspotted Razvan from ordinary Razvan—that is, between their snapshots—is comparable to the difference between pre-incident Gage and post-incident Gage, at least with respect to being sensitive to reasons regarding fidelity. Thus, since the non-occurrence of the smiling and the tilting is what would make the two snapshots sufficiently different, my proposal entails that we hold fixed the nonoccurrence of any other action from his spouse other than the smiling and the tilting of her head.

This move carries over to Trauma. The snapshot of Katharina in the actual world is considerably different from her snapshot if she were confronted with someone other than David. This difference is caused by the presence of David. So, we fix the nonoccurrence of any other person appearing before Katharina except David.

The result with my proposal is this. In a sufficient proportion of nearby worlds, Razvan's spouse will be smiling while tilting her head in a partic-

ular way to prevent Razvan from seeing sufficient reasons pertaining to the fidelity of his spouse. As such, Razvan will possess the correct modal properties for agent-based modalists to say that he is not reasons-sensitive. Similarly, Katharina, in a sufficient proportion of worlds, will be confronted with the presence of David, who deactivates her trauma, allowing her to translate reasons for and against verbally abusing people into action, which in turn makes her possess the right modal properties for agent-based modalists to call her reasons-sensitive.

At this point, one might argue that the agent-based view risks being overly flexible.¹⁹ In particular, the concern is that the notion of an agent-altering event can be selectively applied to preserve the desired verdicts. In cases like Jealousy, I claim that the presence or absence of a specific trigger—such as the spouse’s smiling and tilting—constitutes an agent-altering difference that justifies excluding nearby worlds from the modal profile. But why isn’t the same move available in more ordinary cases? Suppose a non-addict refrains from taking a drug in the actual world when given a reason not to, but takes it in a nearby world in which that reason is absent. Why not say that the absence of that reason is itself agent-altering, and thereby exclude those worlds? What exactly distinguishes Razvan from the non-addict? Without a principled account of which differences qualify as agent-altering and which do not, the view risks appearing ad hoc.

¹⁹I thank an anonymous reviewer for raising this worry.

In my view, what matters in considering whether an event is agent-altering or not is whether it fundamentally changes what one values and how one exercises their mental faculties. As I see it, the presence of a reason to refrain from drug use—while motivationally significant—is not typically the kind of difference that alters the agent’s evaluative and deliberative structure deeply across a sufficient proportion of worlds. To press this point: in the non-addict case, the counterfactual snapshot remains deeply continuous with the actual one. It still reflects the same evaluative system, albeit in a context where a specific consideration is missing. In that sense, the non-addict in the counterfactual world is still recognizably the same agent. Contrast this with Razvan, or with tumor-afflicted Whitman, where the divergence involves a structural discontinuity: the snapshot is no longer a genuine counterpart of the agent. There, we see a significant rupture in the agential structure itself, not merely a shift in situational inputs.

5 Conclusion

My aim with this paper was two-fold. The first aim was to dampen the hopes for the mechanism approach. The mechanisms view allows for the existence of composite mechanism, the response to which ultimately leads to an unattractive disjunction for mechanism modalists: Either their view collapses into agent modalism, or their mechanisms lose the flexibility required for reasons-sensitivity. I also argued that dropping the agent

out of the picture (i) requires a conceptual rethinking of our responsibility ascriptions, (ii) detaches the mechanism view from the initial appeal of modalism, and (iii) requires the view to argue for an ownership condition that many find problematic.

The concerns I raised against the mechanism approach do not thereby satisfy the second aim, which was to bring back agent modalism on the table. To do this, in Section 4, I argued that agent modalists should fix the nonoccurrence of agent-altering events to overcome Frankfurt cases. I then showed that this strategy successfully handles rational blind spot cases, reinforcing its plausibility. Whether this approach can meet other challenges to modalism remains an open question.²⁰ Still, it offers a promising way to preserve the agent-centred view of reasons-sensitivity while avoiding the pitfalls of the mechanism view.

²⁰For example, Sartorio (2015) argues that modalism is incompatible with what she calls the ‘Supervenience’ claim elicited by Frankfurt cases. Others, including Sartorio, argue that modalism is threatened by experiments in social psychology (see McKenna and Warmke (forthcoming), Sartorio (2017b), Vargas (2013)).

References

- Bird, A. (1998). Dispositions and antidotes. *Philosophical Quarterly*, 48(191), 227–234.
- Brink, D. O., & Nelkin, D. K. (2013). Fairness and the architecture of responsibility. *Oxford Studies in Agency and Responsibility*, 1, 284–313.
- Choi, S. (2009). The conditional analysis of dispositions and the intrinsic dispositions thesis. *Philosophy and Phenomenological Research*, 78(3), 568–590.
- Cross, T. (2005). What is a disposition? *Synthese*, 144(3), 321–41.
- Fara, M. (2008). Masked abilities and compatibilism. *Mind*, 117(468), 843–865.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility* (M. Ravizza, Ed.). Cambridge University Press.
- Fisher, J. C. (2013). Dispositions, conditionals and auspicious circumstances. *Philosophical Studies*, 164(2), 443–464.
- Frankfurt, H. G. (1969). Moral responsibility and the principle of alternative possibilities. *Journal of Philosophy*, 66(23), 829–839.
- Gearhart, S. (2013). Phineas gage.
- Ginet, C. (2006). Working with fischer and ravizza's account of moral responsibility. *The Journal of Ethics*, 10(3), 229–253.
- Gundersen, L. (2002). In defence of the conditional account of dispositions. *Synthese*, 130(3), 389–411.

- Haji, I. (1998). *Moral appraisability: Puzzles, proposals, and perplexities*. Oxford University Press.
- Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132(1), 109–136.
- Heering, D. (2022a). Actual sequences, frankfurt-cases, and non-accidentality. *Inquiry*, 65(10), 1269–1288.
- Heering, D. (2022b). Reasons-responsiveness, modality and rational blind spots. *Philosophical Studies*, 180(1), 293–316.
- Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98(6), 273–299.
- Kaiserman, A. (2021). Reasons-sensitivity and degrees of free will. *Philosophy and Phenomenological Research*, 103(3), 687–709.
- Lewis, D. (1973). *Counterfactuals*. Harvard University Press.
- Loewenstein, Y. (2023). Reasons-responsiveness, control and the negligence puzzle. *Philosophical Issues*, 33(1), 124–139.
- Malzkorn, W. (2000). Realism, functionalism and the conditional analysis of dispositions. *Philosophical Quarterly*, 50(201), 452–469.
- Manley, D., & Wasserman, R. (2008). On linking dispositions and conditionals [Publisher: Oxford University Press]. *Mind*, 117(465), 59–84.
- McKenna, M. (2000). Assessing reasons - responsive compatibilism. *International Journal of Philosophical Studies*, 8(1), 89–114.
- McKenna, M. (2001). Review of responsibility and control: A theory of moral responsibility. *The Journal of Philosophy*, 98(2), 93–100.

- McKenna, M. (2005). Reasons reactivity and incompatibilist intuitions. *Philosophical Explorations*, 8(2), 131–143.
- McKenna, M. (2013). Reasons-responsiveness, agents, and mechanisms. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility volume 1* (151–183). Oxford University Press.
- McKenna, M., & Warmke, B. (forthcoming). Does situationism threaten free will and moral responsibility? *New Content is Available for Journal of Moral Philosophy*.
- Mumford, S. (1998). *Dispositions*. Clarendon Press.
- Nelkin, D. K. (2011). *Making sense of freedom and responsibility*. Oxford University Press.
- Russell, P. (2002). Responsibility and control: A theory of moral responsibility. *Canadian Journal of Philosophy*, 32(4), 587–606.
- Sartorio, C. (2006). On causing something to happen in a certain way without causing it to happen. *Philosophical Studies*, 129(1), 119–136.
- Sartorio, C. (2015, September). Sensitivity to Reasons and Actual Sequences. In *Oxford Studies in Agency and Responsibility: Volume 3*. Oxford University Press.
- Sartorio, C. (2016). *Causation and free will*. Oxford University Press UK.
- Sartorio, C. (2017a). Frankfurt-style examples. In K. Timpe, M. Griffith, & N. Levy (Eds.), *The routledge companion to free will* (179–190). Routledge.
- Sartorio, C. (2017b). Situations and responsiveness to reasons. *Noûs*, 52(4), 796–807.

- Smith, M. (1997). A theory of freedom and responsibility. In G. Cullity & B. N. Gaut (Eds.), *Ethics and practical reason* (293–317). Oxford University Press.
- Smith, M. (2003). Rational capacities, or: How to distinguish recklessness, weakness, and compulsion. In S. Stroud & C. Tappolet (Eds.), *Weakness of will and practical irrationality* (17–38). Oxford University Press.
- Stout, N. (2016). Reasons-responsiveness and moral responsibility: The case of autism. *The Journal of Ethics*, 20(4), 401–418.
- Todd, P., & Tognazzini, N. A. (2008). A problem for guidance control [Publisher: Wiley-Blackwell]. *Philosophical Quarterly*, 58(233), 685–692.
- Vargas, M. (2013). Situationism and moral responsibility: Free will in fragments. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), *Decomposing the will*. Oxford University Press USA.
- Vihvelin, K. (2004). Free will demystified: A dispositional account. *Philosophical Topics*, 32(1-2), 427–450.
- Vihvelin, K. (2013). *Causes, laws, and free will: Why determinism doesn't matter*. Oxford University Press.
- Wallenfeldt, J. (2016). Texas tower shooting of 1966.
- Watson, G. (2001). Reasons and Responsibility [Publisher: The University of Chicago Press]. *Ethics*, 111(2), 374–394.
- Yablo, S. (2002). De facto dependence. *Journal of Philosophy*, 99(3), 130–148.

Yaffe, G. (2000). Review of john fischer and mark ravizza's responsibility and control: A theory of moral responsibility. *Erkenntnis*, 53(3), 429–434.

Syracuse University, USA