ABSTRACT:

I argue for responsibility internalism. That is, moral responsibility (i.e., accountability, or being apt for praise or blame) depends only on factors internal to agents. Employing this view, I also argue that no one is responsible for what AI does but this isn't morally problematic in a way that counts against developing or using AI.

Responsibility is grounded in three potential conditions: the control (or freedom) condition, the epistemic (or awareness) condition, and the causal responsibility condition (or consequences). I argue that causal responsibility is irrelevant for moral responsibility, and that the control condition and the epistemic condition depend only on factors internal to agents. Moreover, since what AI does is at best a consequence of our actions, and the consequences of our actions are irrelevant to our responsibility, no one is responsible for what AI does. That is, the so-called responsibility gap exits. However, this isn't morally worrisome for developing or using AI. Firstly, I argue, current AI doesn't generate a new kind of concern about responsibility that the older technologies don't. Then, I argue that responsibility gap is not worrisome because neither responsibility gap, nor my argument for its existence, entails that no one can be justly punished, held accountable, or incurs duties in reparations when AI causes a harm.

RESPONSIBILITY INTERNALISM & RESPONSIBILITY FOR AI

by

Huzeyfe Demirtas

B.A., Firat University, 2009

M.A., Syracuse University, 2022

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Philosophy.

Syracuse University

May 2023

ACKNOWLEDGEMENTS:

I would like to begin by profusely thanking my advisor, Ben Bradley. I am still amazed at how lucky I am to have known him and learned from him. It has been an absolute privilege and an honor to be his student. He is the philosopher, teacher, and human being I can only aspire to be.

My committee members: Sara Bernstein, Mark Heller, and Hille Paakkunainen. Sara's work generated and inspired my entire dissertation project. Mark has never spared his most difficult questions. Hille is pretty much the reason why I started doing moral philosophy. It's something close to a lucky cosmic accident for me to have had them around to help me over the years as I wrote my dissertation.

Two people deserve thanks at least as much as my committee members: Stephen Kershnar and Andrew Khoury. Conversations with Steve, his work, and his ever judicious and clear comments on almost every chapter of my dissertation have been immensely helpful. The insights I found in Andy's work have been a life saver at many crucial points in my dissertation.

I have also been fortunate enough to benefit from many excellent friends and philosophers through conversations or their written comments: Andrei Buckareff, Ben Cook, Neil Feit, Kellan Head, Johannes Himmelreich, Jessica Isserow, Vera Hoffmann-Kolss, Max Kistler, Sanggu Lee, Yaojun Lu, Kris McDaniel, Matthias Rolffs, Carolina Sartorio, Byron Simmons, Jags Singh, Jan Swiderski, and Joshua Tignor. Many, many thanks to every one of them!

I have been blessed with a wonderfully supportive family: my dad, Vahdettin Demirtas, my mom, Hidayet Demirtas, and my brother, Talha Demirtas. They have always humbled me with their love and kindness, their faith in me, and by taking pride in my work. Thanks to my friend, Mehmet Berk, for igniting my interest in philosophy long years ago and for all the philosophy conversations over the years. Thanks to my friend, Anatole Ruslanov, for his valuable guidance, for helping me to get back to academia, and eventually to start my graduate studies. And, indeed, a special thanks to Lamyae Kerzazi for all the intellectually stimulating conversations for the last several years, for being the most kind and understanding partner, and a best friend.

Not to forget—thanks to ChatGPT! I ran some of the sentences above in this page by it. I suppose this is befitting of a dissertation that is also about AI. (Of course, some readers will wonder if I can square this gratitude to AI with some of my theses in the last chapter.)

Contents

Introduction

I will argue that responsibility internalism is true. Responsibility internalism is the idea that moral responsibility depends only on factors internal to agents. Before I close, I will also show how responsibility internalism weighs in on an important debate about responsibility in the context of artificial intelligence—namely, the debate over 'responsibility gap.'

In this introduction, I further clarify the above claims and explain how my arguments will unfold. Let's start by clarifying "responsibility." By "S is *responsible for* Φ," we might mean at least three different things. First, we might mean that S *caused* Φ. Second, we might mean that S *has a duty* regarding Φ. And third, we might mean that S is apt for moral *blame or praise for* Φ. Responsibility internalism is a thesis about the last one. Let's now further elaborate on each of these.

I will call the first one "causal responsibility." Broadly speaking, causal responsibility occurs when something or someone is in the causal chain that generated an effect. That is, that which is seen as a cause of an effect by the correct theory of causation is causally responsible for that effect. To illustrate, suppose Suzy pushes the main door open to a building and thereby hits Billy, who has been idly standing on the other side of the door, causing him to fall and get hurt. Plausibly, although Suzy is not to be blamed for Billy's getting hurt, she's causally responsible for it. But notice that she's not causally responsible only for Billy's getting hurt. She's also causally responsible for pushing the door, and for moving her arm to push the door. Moreover, the causal chain that ended up generating Billy's getting hurt involves mental

elements such as deciding, and trying, setting out, or willing to push the door. So, causal responsibility involves both mental and extramental elements. For reasons that will be clearer below, I reserve the term causal responsibility to involve only the extramental elements.

The second thing we might mean by "S is *responsible for* Φ" is that S has a duty regarding Φ. Hence, I will call this the "duty sense of responsibility." This sense of responsibility simply refers to one's moral obligations. If S is a teacher, S is responsible for imparting knowledge to her students—i.e., it's S's duty or obligation to teach them. Or overall, it might be everyone's moral duty to maximize the good outcomes, act accordingly with the Categorical Imperative, or do what a virtuous person would do in the circumstances. Although the duty sense of responsibility will take part in our discussion below, responsibility internalism is not about this sense of responsibility. Hence, I will not use "moral responsibility" to refer to the duty sense of responsibility unless I make it explicit.

The third thing we might mean by "S is *responsible for* Φ" is that S is apt for blame or praise for Φ. As it's often called in the literature, I will call this the "basic desert responsibility" (Pereboom 2014:2, Sartorio 2016:7). *Desert* in the sense that if S is responsible for Φ, S deserves blame, resentment, or punishment, or credit, gratitude, praise, or reward, for Φ. *Basic* in the sense that if S is to blame or praise for Φ, S is to blame or praise because of Φ and not because of any instrumental value that blaming or praising S might bring about. The blame or praise in question might be moral or non-moral. Responsibility internalism is only about the former—*moral* praise or blame. This sense of responsibility might further be divided into two categories—*synchronic* responsibility and *diachronic* responsibility (Khoury 2013, Khoury and Matheson 2018). While synchronic responsibility concerns S's responsibility at the time of

action, diachronic responsibility concerns S's responsibility at a later time. The former doesn't entail the latter. While S might deserve blame for Φ now, she might deserve no blame for Φ after much repentance at a later time. I will take responsibility internalism to be only about synchronic responsibility, but much of what I'll say applies also to diachronic responsibility.

Some philosophers might insist that I say more to clarify the relevant sense of moral responsibility. It used to be that philosophers writing on moral responsibility had (or seemed to have) in mind roughly the same thing—though it's easier to say what they *didn't* have in mind than to say what they *did* have in mind. They thought it's neither causal responsibility nor *having* moral responsibilities or obligations that they have in mind. But whatever they might have had in mind, the literature now involves various 'distinct' conceptions of moral responsibility: accountability, attributability, and answerability. Accountability is the same as basic desert responsibility. According to attributability conception of responsibility, to be morally responsible for something is to be such that that thing is properly attributed to one. According to answerability conception of responsibility, to be morally responsible for something is to be in principle an appropriate target of demands for justificatory reasons for that thing.

Some philosophers hold that accountability, attributability, and answerability refer to three distinct types of moral responsibility (Shoemaker 2011, 2015). Some hold that there is only one type of moral responsibility, and it is best captured by answerability (Hieronymi 2014, Smith 2012, 2015). Some hold that attributability is the one real conception of moral

responsibility.[1] Some hold that attributability and answerability are not really conceptions of moral responsibility (Levy 2005).[2] Some note these controversies as among the reasons for a methodological 'morass' in the responsibility literature (Shoemaker 2020). And finally, some even note all these controversies among the reasons that responsibility studies should divide into distinct subfields (Rudy-Hiller 2021). I take no position on these controversies. Although much of what I will say apply also to attributability and answerability, I take responsibility internalism to be about accountability.

Responsibility internalism then is the idea that basic desert responsibility depends only on factors internal to agents. So, we have a few more things to clarify here: 'depends,' 'factors,' 'internal,' and 'agents.' By "depends," I mean metaphysically and not, say, epistemologically. That is, that which grounds or explains moral responsibility, or that which is the responsible-maker. By "agents," I have in mind individuals and not collective agents or groups such as corporations, peoples, and governments. By "internal," I mean that which is not external. For instance, causal responsibility, as defined above, is or at least partly constituted by that which is external to agents and hence is not internal to them. Below when I present how my argument for responsibility internalism will unfold, I'll further clarify what I have in mind by "internal" and by "factors."

[1] Some philosophers (like Hieronymi, and Smith) who defend answerability are sometimes described by others as among those who defend attributability. This is partly due to the debate over whether answerability is distinct from attributability. See Talbert (2019), sections 3.1.2 and 3.1.3, for a brief survey of defenders of attributability or answerability.

[2] More precisely, Levy argues that attributability is not a kind of moral responsibility. But he also argues that—what Smith calls—'responsibility as answerability' only provides conditions for attributability. So, his argument targets both attributability and answerability.

Here's my argument for responsibility internalism in a nutshell. In the literature, we find basically three potential conditions for moral responsibility: the control (or freedom) condition, the epistemic (or awareness) condition, and the causal responsibility condition (or consequences).[3] I will argue that causal responsibility is metaphysically irrelevant for moral responsibility, and that the control condition and the epistemic condition depend only on factors internal to agents.

I don't claim that everyone in the literature accepts that all these three factors are metaphysically relevant for moral responsibility. For instance, some philosophers don't think that consequences affect moral responsibility, and I will side with these philosophers myself. So, all these three factors are not always found together in the literature. My claim rather is that, to a first approximation, the list of what various philosophers think as the fundamental factors that affect or ground moral responsibility consists of these three factors.

I already clarified causal responsibility above. So, let me briefly clarify the other two conditions: control and awareness. Consider again the case above. Suzy pushes the main door to a building open and thereby hits Billy, who has been idly standing on the other side of the door, causing him to fall and get hurt. Notice that if Suzy didn't freely do what she did—i.e., if she wasn't in control of her action—she isn't blameworthy for hurting Billy. For instance, if Suzy were under hypnosis when she pushed the door, plausibly she isn't blameworthy for what she

[3] See, e.g., Cyr (2019:479), Fischer and Ravizza (1998) (see ch.4 for the discussion on responsibility for consequences), Haji (2008:17), Hartman (forthcoming), Kaiserman (2018:5), Kaiserman (2021:3602-3), Khoury and Matheson (2018:205), Moore (2011:496), Mumford (2013:109-10), Pereboom (2014:2), Rosen (2011:405), Rudy-Hiller (2018), Sartorio (2007) (especially, pp.750-1), Sartorio (2016:7-8), Talbert (2019), Vallentyne (2008:62), Van Inwagen (2015:283), Vincent (2011) (especially, pp.19-21), Wieland (2017:3-5), Wolf (1987), Zimmerman (1997:410-1).

did. As this brief illustration also suggests, the control condition is basically the subject of classic free will debate. However, even if Suzy freely did what she did, it doesn't yet follow that she's blameworthy. If she weren't aware, or couldn't reasonably have been expected to be aware, that Billy was standing on the other side of the door, she is again not blameworthy for what she did. So then being in control of one's action and being aware of various factors that are relevant to one's action are necessary for moral responsibility. I'll say more about both these conditions in the following chapters. But this should suffice for now.

I said above that 'to a first approximation' it is these three factors in question that can affect or ground moral responsibility. This is because I think there are other factors, such as one's motivation, intention, or care in performing an action, that affect moral responsibility. Consider Kant's famous example of the prudent shop keeper who never overcharges his customers even if they're children. Yet he does this not because it's the right thing to do but because it's good for his business. He doesn't want a bad reputation as a business owner. Assume that he knows that he's doing the right thing despite the fact that he's not motivated by the rightness of his action. Rather what motivates him is his own self-interest. Assume also that he is in control of his own action. So, he satisfies all three potential conditions for responsibility. Now we don't necessarily need to think that he's blameworthy because he's not motivated by the rightness of his action. But it seems plausible to think at least that he would be *more* praiseworthy if he were 'motivated' by the rightness of his action instead, or if his 'intention' were to do the right thing as he knew it rather than to serve his own self-interest, or if he 'cared' more about the rightness of his action. It follows that motivations, intentions, or cares also affect moral responsibility.

I'm not sure if the motivation, intention, or care that are at stake in the above case are all distinct from one another. It might be that they all ultimately pick out just one factor that's relevant for responsibility. I don't take a position on this. And although I'll sometimes appeal to them in my discussion below, I will largely exclude them from my argument for responsibility internalism. This is for two reasons. One, motivations, intentions, or cares all seem to be factors internal to agents. And if they are, they are no threat to responsibility internalism. Two, if one has reservations that they are solely internal factors, one could appeal to appropriate versions of pretty much all I will say in favor of the epistemic condition being solely an internal concern.

Here are the details of how the argument for responsibility internalism will unfold. Some philosophers hold that, all else equal, one's degree of moral responsibility is proportionate to one's degree of causation (or causal contribution). Call this thesis **Proportionality**. If causation doesn't come in degrees, **Proportionality** is false or uninteresting. So, in **chapter one**, I discuss whether causation comes in degrees. I argue that it does by showing that all the main objections in the literature against graded causation fail and that denying graded causation is theoretically too costly. This chapter is a slightly revised version of my paper '*Causation Comes in Degrees'* that's published in *Synthese* (Demirtas, 2022b).

In **chapter two**, I argue that **Proportionality** is false despite the fact that causation comes in degrees. To establish this, I employ six different criteria for measuring degrees of causation and show that **Proportionality** understood according to each of these criteria entails implausible results. I also show that we don't need **Proportionality** in order to account for the kind of cases that motivate **Proportionality**. This chapter is a slightly revised version of my

paper '*Moral Responsibility Is Not Proportionate to Causal Responsibility'* that's published in *The Southern Journal of Philosophy* (Demirtas, 2022c).

In **chapter three**, I argue that there is no resultant moral luck. What's at stake in the debate over resultant moral luck is best cast in terms of whether causal responsibility increases one's degree of moral responsibility. And the proponents of resultant moral luck hold that it does. I draw attention to previously unexplored implications of resultant moral luck and argue that these implications leave resultant moral luck more indefensible than it was thought to be. I also show that what's typically taken to be the positive argument in favor of resultant moral luck fails. I conclude that we should reject resultant moral luck. This chapter is a slightly revised version of my paper '*Against Resultant Moral Luck'* that's published in *Ratio* (Demirtas, 2022a).

In **chapter four**, I argue that causal responsibility is metaphysically irrelevant for moral responsibility. This is because causal responsibility figures in explaining neither the *degree* of moral responsibility nor what one is morally responsible *for* (i.e., what *makes* one morally responsible). I also defend this view against potential objections and show various theoretical advantages of it.

In **chapter five**, I argue that neither the epistemic condition nor the control condition presupposes anything external to agents. The epistemic condition rests on the idea, roughly, that one can be morally responsible only if one is aware of certain morally relevant factors. The awareness in question can be knowledge, justified true belief, true belief, or belief. As it is commonly accepted, knowledge is too strong a requirement for moral responsibility. I follow the reasoning behind this and show that *justified* true belief is also too strong a requirement. I

further argue that moral responsibility doesn't require even *true* belief. And since the awareness requirement in question presupposes no form of *true* belief, it doesn't presuppose anything external to agents.

The control condition is the subject matter of the classic free will debate. I survey the leading compatibilist and incompatibilist theories of control and argue that none of them, at least in their most plausible forms, presupposes anything external to agents. A major concern for my argument is that the debate between compatibilists and incompatibilists mainly revolve around determinism. Compatibilists argue that the kind of control required for moral responsibility—i.e., free will—is compatible with determinism, and incompatibilists reject this. Determinism is the idea that at any moment the state of world and the laws of nature entail one unique future. As it stands, determinism is not only a feature internal to agents but a feature of the world. However, I argue, (in)determinism external to agents is irrelevant for the control condition—what matters is only (in)determinism internal to agents. That is, what matters is only whether the mental states of agents are (un)determined, not whether anything else in the universe is.

I conclude that the epistemic condition and the control condition depend only on factors internal to agents. Since causal responsibility is also irrelevant to moral responsibility, there remains no condition for moral responsibility that depends on anything external to agents. Hence, responsibility internalism is true.

In **chapter six**, I employ responsibility internalism to weigh in on a debate about responsibility in the context of artificial intelligence. Consider autonomous systems or machines

that rely on artificial intelligence such as self-driving cars, lethal autonomous weapons, candidate screening tools, medical systems that diagnose cancer, and automated content moderators. Who is responsible for it when such machines or systems (or AI for short) causes a harm? Given that current AI is far from being conscious or sentient, it is unclear that AI is responsible for a harm it causes. But given that AI gathers new information and acts autonomously, it is also unclear that those who develop or deploy AI are responsible for what AI does. This leads to the so-called responsibility gap: that is, roughly, cases where AI causes a harm, but no one is responsible for it. Two central questions in the literature are whether responsibility gap exists, and if yes, whether it's morally problematic in a way that counts against developing or using AI. While some authors argue that responsibility gap exists, and it is morally problematic, some argue that it doesn't exist or that it's dubious that it exists. Drawing from discussions in the earlier chapters, I defend a novel position. I firstly argue that current AI doesn't generate a new kind of concern about responsibility that the older technologies don't. Then, I argue that responsibility gap exists—that, more precisely, responsibility gap is inevitable and ubiquitous—but this is not morally problematic in a way that counts against developing or using AI.

Chapter 1: Causation Comes in Degrees

Abstract:

Which country, politician, or policy is more of a cause of the Covid-19 pandemic death toll? Which of the two factories causally contributed more to the pollution of the nearby river? A wide-ranging portion of our everyday thought and talk, and attitudes rely on a graded notion of causation. However, it is sometimes highlighted that on most contemporary accounts, causation is on-off. Some philosophers further question the legitimacy of talk of degrees of causation and suggest that we avoid it. Some hold that the notion of degrees of causation is an illusion. In this chapter, I'll argue that causation does come in degrees.

Which is the main cause of obesity—genetics, or unhealthy diet? Was it due more to Suzy's absence or the lack of good food that last night's party was a bore? Which of the two factories contributed more to the pollution of the nearby river? Is alcohol more important cause of heart attack than smoking? We typically don't question the legitimacy of ordinary questions like these. Yet these questions might suggest something controversial—that causation admits of degrees.

It's sometimes highlighted that on most contemporary accounts causation is on-off (Sartorio 2010, 2015a). Some authors question whether the talk of degrees of causation is legitimate (Pearson 1980, Barker and Steele 2015). Some hold that the talk of degrees of causation is misleading (Zimmerman 1985). It's also argued that the notion of degrees of causation is an illusion and needs to be explained away (Sartorio 2020). As opposed to this,

some authors hold the thesis that causation comes in degrees.[4] However, the literature on this thesis has two big gaps. One, it's often unclear what's meant by degrees of causation. And two, most importantly, the main objections against this thesis remain unanswered. Hence, it would seem, this thesis hasn't properly been defended in the literature. This chapter aims to fill these gaps and argues that causation comes in degrees.[5]

In **§1**, I'll distinguish four different senses of degrees of causation and clarify the thesis that I'll argue for. In **§2** through **§4**, I'll list the three main objections from the literature to the thesis that causation comes in degrees and argue that they fail. In **§5**, I'll argue that it's theoretically too costly to deny this thesis. I'll conclude—since we have no good reason to reject this thesis and it's too costly to deny it—that this thesis is true.

1. Four Senses of Degrees of Causation

In this section, I'll distinguish four different senses of degrees of causation in the literature. First, suppose Suzy and Billy are both very funny. Their attendance to tonight's office party made it an exceptionally fun party. We might wonder about Suzy's causal significance on the party's being fun compared to Billy's causal significance. That is, for the causal factors c_1, c_2, c_3,… c_n, in the previous causal link that brought e about, we might wonder about the relative causal influences of c_1, c_2, c_3,… c_n on e. For instance, the degree of Billy's causal significance

[4] Cf., e.g., Northcott (2005b), Braham and van Hees (2009), Moore (2011, 2013), Mumford (2013), Bernstein (2016, 2017), and Sprenger (2018).

[5] To be clear, I'll defend the thesis that there are degrees of causation, and not any specific account of (measuring) degrees of causation. Also, one might object that Moore (2009) defends this thesis. Indeed, in chapter 3 (p.71) of his book he says that he'll defend this thesis in chapters 5-6. However, in chapters 5-6 he says no more in favor of this thesis than, roughly, that it is assumed in legal practice. As Bebe puts in her review of Moore's book, "Moore offers very little by way of motivation or argument" for this thesis (Bebe 2013, p.105).

might be twice the degree of Suzy's causal significance. After Northcott (2005b), I'll call this the *relative* sense of degrees of causation. Degrees of causation in this sense is endorsed by most, if not actually all, those who hold that causation comes in degrees.[6] And I'll argue that causation comes in degrees in this sense.

Notice that the relative sense of degrees of causation could be applied not only within a case but also across distinct cases. Imagine that there was another and equally fun office party in the next building tonight. Imagine also that Timmy and Lily were the funny people in this second party. We might wonder about, say, Billy's causal significance to the first party's being fun relative to Lily's causal significance to the second party's being fun. It might be that the degree of Lily's contribution to the second party's being fun is twice the degree of Billy's contribution to the first party's being fun. This might well be the case if Lily is two times funnier than Billy.[7]

Notice also that the examples above are instances of token-level causal relations. We can also consider type-level causal relations—e.g., alcohol causes heart attack. One might, for instance, hold that alcohol is a more significant cause of heart attack than smoking because, say, in more individual instances of heart attacks alcohol is the more significant cause than

[6] Kaiserman (2016) distinguishes between scalarity of causal relations and degrees of causal contribution, and rejects the former and endorses the latter. His notion of scalarity of causal relations will, in my taxonomy, fall under the fourth sense of degrees of causation below. His notion of degrees of causal contribution is basically what I call the relative sense of degrees of causation. "Degrees of causation," "degrees of causal contribution," and their variations are often used interchangeably in the literature both by those who think that causation comes in degrees and those who don't (cf. e.g., Moore (2009, p.71), Braham and van Hees (2009, p.324), Sartorio (2010, p.836), Alexander and Ferzan (2012, p.83), Mumford (2013, p.109)).

[7] Among those who hold that causation comes in degrees, Moore (2012, p.448) might be pessimistic about relative sense of degrees of causation across distinct cases. However, it's not clear enough whether what he has in mind is this one or the third sense of degrees of causation I'll mention below.

smoking or because alcohol causes heart attacks more often than smoking. Objections in the literature against degrees of causation typically focus only on token-level causal relations.[8] Since, as I'll argue, these objections fail, they fail for both types of causal-relations. Moreover, as the discussion below will further demonstrate, it's just as costly to deny degrees of causation in type-level relations. Hence, I'll be arguing for degrees of causation in both sorts of causal relations.

Second, suppose Suzy, Billy and all their colleagues received the good news earlier that day that they all got a raise. So, they all were already in a good mood, which was among the reasons why tonight's party was fun. We might now wonder about Billy's causal impact on the party's being fun compared to the causal impact of the good news. That is, given an outcome *e*, we might wonder about the causal strength of a cause of *e* in the previous link compared to a cause of *e* in a link further back in the causal chain. This sense of degrees of causation is essentially no different than the relative sense of degrees of causation. But I must qualify this. It's typically accepted that causation is transitive: roughly, if *d* causes *c*, which in turn causes *e*, then *d* causes *e*. Yet some think that causation is *in*transitive. If causation is intransitive, in some cases there may be no degrees of causation in this second sense. This is because if *d* is not a cause of *e*, there's no question of the extent to which *d* causes *e*. However, that causation is intransitive doesn't entail that it's *never* the case that if *d* causes *c*, and *c* causes *e*, then *d*

[8] I suspect this is because it seems easier to argue for degrees of causation in type-level relations—especially once degrees of causation in token-level relations is established.

causes *e*. Transitivity might still sometimes obtain.[9] Hence, even if causation is intransitive, there may still be degrees of causation in this second sense—only in a more limited scope.

The debate in the literature over this sense of degrees of causation—rather than being directly on whether causation comes in degrees in this sense—is typically conducted via a specific criterion for measuring degrees of causation.[10] The criterion in question, roughly, is that causal strength gets diluted as it's transmitted from one causal link to another. For instance, a causal factor, *d*, that's further back in the causal chain might have smaller causal influence on an outcome, *e*, compared to another cause, c_1 *or* c_2, of *e* that's in the link right before *e*. Indeed, that might depend on the initial causal strength of *d* and how many causal links it's transmitted through before *e*. If *d* had a massive causal influence to begin with, even its diluted causal influence might turn out to be bigger than the causal influence of c_1 *or* c_2 (Moore 2013, p.127). Now notice that the criterion in question is about how to measure the relative strength of causal factor, *d*, that's further back in the causal chain. But even if this criterion is false, it doesn't follow that the causal influence of *d* cannot be compared to the causal influence of c_1 *or* c_2. If *d is* a cause of *e*, I see no reason why we can compare the causal influence of c_1 to that of c_2, but can't compare the causal significance of *d* to that of c_1 *or* c_2. That is, if causation comes in degrees in the first sense above, it also comes in degrees in this second sense. As Moore—the biggest defender of the criterion above—also puts, of the two senses of degrees of

[9] Cf. Hall (2000), and Paul and Hall (2013, ch.5) for further discussion on transitivity of causation. Cf. Mumford (2013) who argues that his account of causation entails that causation is *not always* transitive.
[10] Cf. e.g., Moore (2009, pp.121-3), Alexander and Ferzan (2012, p.84), Mumford (2013, p.111), Bebe (2013, pp.105-6), Tadros (2018, pp.428-9).

causation above, the first one is the most basic (Moore 2013, p.123). Hence, I'll focus on degrees of causation in the first sense above.

Third, in addition to Suzy's impact on the party's being fun compared to, say, Billy's impact, we might also wonder about Suzy's impact in its own right. For instance, given that everyone in the office is already in a good mood, we might wonder what the impact of Suzy's presence or absence would be on the party's being fun. After Northcott (2005b), I'll call this the *absolute* sense of degrees of causation. That causation comes in degrees in this sense intuitively seems plausible. However, among those who believe that causation comes in degrees, it's unclear whether most of them endorse degrees of causation in this sense.[11] The kinds of examples they use to illustrate their views don't strongly suggest an endorsement or rejection of this sense of degrees of causation. Moreover, among those who reject that causation comes in degrees, I'm not aware of anyone who objects particularly to this sense of degrees of causation. But much of what I say below works also in favor of degrees of causation in this sense.

Fourth, it's commonly held that causation is a relation between cause and effect. One might then wonder whether the idea that causation comes in degrees might mean that an effect can be more caused or less caused. I'll call this—the idea that an effect can be more/less caused—the *scalarity of effect* sense of degrees of causation. Considering again that tonight's

[11] However, cf. Northcott (2005b) for a defense of this sense of degrees of causation. Cf. Fitelson and Hitchcock (2011) for a survey of various measures of degrees of causation some among which can plausibly be considered as aiming to measure degrees of causation in this sense. Also, one way to see the difference between the relative sense and the absolute sense of degrees of causation is to think of the former as an ordinal measure and the latter as a cardinal measure. However, the word "absolute" can still be a bit misleading. One might, for instance, think that it implies that a given causal factor has a fixed quantity of causal influence that applies in all contexts and for all outcomes. This seems wrong. Cf. Northcott (2008, pp.88-9) for a related discussion.

party was fun, the question is whether it could have been more caused or less caused. Or supposing that it's raining outside now, the question is whether it could be said to be more caused or less caused. Among those who believe that causation comes in degrees, some explicitly reject it in the scalarity of effect sense.[12] For most others, their own illustrations of what they have in mind and their context strongly suggest that scalarity of effect sense is not in their purview. Among those who reject degrees of causation, I'm not aware of anyone raising an objection explicitly to this sense of degrees of causation. Hence, I'll leave this sense of degrees of causation aside.

2. Objection 1: On Most Contemporary Accounts Causation is On-Off

As a worry about the thesis that causation comes in degrees, it is sometimes noted that on most contemporary accounts causation is all or nothing, that it doesn't admit of degrees.[13]

However, what exactly should be worrisome is unclear. This is because it's unclear how the accounts in question should suggest that causation doesn't admit of degrees. The worry might arise because most accounts of causation "focus on the binary question of whether something is a cause" (Lagnado, Gerstenberg, Zultan 2014). One might then be tempted to think that something either is or isn't a cause of an outcome—there are no other options.

[12] Moore calls the idea that some event can be more or less caused "bizarre" (2012, p.447). Kaiserman argues that causation is "not a scalar relation" (2016, p.389). It seems to me from his discussion that it is, what I call, the scalarity of effect sense of degrees of causation that he rejects. Moore sometimes expresses his view of degrees of causation as "causation is a scalar relation" (2009, p.105). But when he further explains his view, it's clear that he has in mind the first two senses of degrees of causation above (2012, pp.446-7).

[13] Cf. Sartorio (2010, 2015a). Mumford (2013), who argues that his account of causation leaves room for degrees of causation, also notes this worry for many other theories of causation.

Hence, causation doesn't come in degrees.[14] But this doesn't follow. That something either is or isn't a cause is just a consequence of the law of excluded middle.[15] It doesn't say anything about whether something could be more or less of a cause. Similarly, it's true that a belief either is or isn't justified. But it doesn't follow that epistemic justification doesn't come in degrees. It does—as it's commonly accepted, a belief can be more or less justified (Feldman 2003, p.21, Pappas 2017).

To further see that it's unclear what should be worrisome, let's take a closer look at some contemporary accounts of causation. For instance, productive theories of causation hold that causation is a matter of producing another event.[16] According to one prominent example of such an account, *c* causes *e* if there's a transfer of conserved quantity (e.g., force, or energy) from *c* to *e* (Dowe 2000). Suppose you hit a glass and it falls off the table. You brought about this outcome in virtue of transferring energy to the glass. Given this picture of causation, one might think that a factor either does transfer energy to an outcome or it doesn't. True enough, but one might still wonder as to how much energy a factor transfers to an outcome. For instance, suppose three people carry a heavy marble top, rectangular dining table from the truck all the way into the kitchen. Two of them, who are regular people, each hold two separate corners on one side, and one of them, who is unusually large and strong, holds both corners on the opposite side. It seems natural to think that the strong one causally contributed more to

[14] Compare: "Causation […] exists or it does not, and if it does exist one does not speak of "degrees" of causation" (Pearson 1980, p.346). Here, Pearson is denying "comparative [causal] contribution." It's clear from his context that what he has in mind is, what I called, the relative sense of degrees of causation.

[15] Thanks to Ben Bradley for the helpful suggestion here.

[16] The question of causal relata is controversial. Events, facts, agents, states of affairs, and features are among the possible candidates. The standard view holds that they're events (Paul and Hall 2013, Schaffer 2016). I don't take a position on this. However, for ease of exposition, I'll treat causation as a relation between events.

the outcome than the other two people. One plausible way to account for this thought is to think that the strong person is more of a cause of the outcome *because* she transferred more energy to the outcome. One might then hold that the degree of causation is about how much energy a given factor transfers to an outcome.[17] Hence, it seems plausible to hold that productive theories of causation can account for degrees of causation.

Let's now turn to counterfactual theories of causation according to which causation is a matter of counterfactual dependence between wholly distinct events (Hall 2004). On the simplistic counterfactual view *c* is a cause of *e* just in case *e* counterfactually depends on *c*. That is, *c* causes *e* just in case had *c* not occurred, *e* would not have occurred. The dependence relation in question doesn't initially seem to leave room for gradation: either the counterfactual dependence holds, or it doesn't. Hence, one might be inclined to hold that counterfactual theories of causation cannot account for degrees of causation. But this is mistaken. Consider the following scenario. Billy and Suzy work in an apple orchard. Today, they're asked to pick up 1000 apples that are ripe, firm and with no indents or discoloration. At the end of the day, Billy counts 400 apples in his basket, and Suzy 600. They bring the apples to the office and call it a day. It is natural to think that Suzy causally contributed more to, or she's more of a cause of, the outcome than Billy. Moreover, it seems plausible to think that Suzy contributed more even if Suzy and Billy spent equal amount of effort. They both worked nonstop, they are equally tired in the end, they spent equal amount of energy and so on. After all, Suzy might just have worked smart rather than hard, and hence got more work done. Now both Billy and Suzy together

[17] Cf. Moore (2012) and Bernstein (2017) for similar suggestions.

stand in a counterfactual dependence relation with the outcome—i.e., if not for both their conducts, the outcome wouldn't have occurred. But it is also true that the portion of the outcome that depends on Suzy's conduct is bigger than the portion of the outcome that depends on Billy's conduct. One might then hold that the degree of causation is about the portion of an outcome that counterfactually depends on a given factor.[18] Hence, it seems plausible to hold that counterfactual theories of causation can account for degrees of causation.

Lastly, let's consider probabilistic theories of causation according to which causation is a matter of raising the probability of the occurrence of an event (Eells 1991). One might again think that a factor either does or it doesn't raise the probability of an outcome. But one might also wonder about how much a given factor increases the probability of an outcome. Consider the following scenario. Suzy had her six-monthly eye exam appointment yesterday, but missed it. She barely feels a problem with her glasses now, but they seem to work fine overall. And since her prescription didn't change in the last two years, she's not too worried about missing the appointment. She needs to drive thirty minutes to work but there's also a heavy rainstorm

[18] This account is inspired by a suggestion made to me by Hille Paakkunainen. A similar suggestion can also be found in Tadros (2018). Notice that the outcome in question is easily decomposable into parts—i.e., 400 plus 600 apples. However, consider multiple people killing someone. It might be difficult to see how the talk of portions of the outcome, i.e., death, will apply in such a case. I'll assume that the general idea here could be adjusted after more sophisticated accounts of the portion of an outcome. There are also more fully developed counterfactual measures of degrees of causation that don't run into this problem. Cf. Chockler and Halpern (2004), and Lewis (2004). Chockler and Halpern's measure receives empirical support from people's actual causal judgements (Gerstenberg et al., 2018, Lagenhoff et al., forthcoming). Also consider pluralists (who hold that there are multiple fundamental concepts of causation) (Hall 2004) and—what might be called—unifiers (who hold that the competing conceptions of causation can consistently be brought together under a single account) (Gerstenberg et al., forthcoming). It's open to both camps to hold that the binary question of whether *c* is a cause of *e* is to be evaluated by one conception of causation (e.g., the counterfactual conception) and the extent of *c*'s causal influence on *e* is to be evaluated by another (e.g., the productive conception, or the probabilistic conception). Such a tiered account can also avoid running into this problem. (Thanks to Matthias Rolffs for the suggestion.)

outside. She's a good driver, and she has driven in rainstorms before. Unfortunately, however, this time around she gets into an accident on her way to work. It seems that the heavy rainstorm is more of a cause of the accident than Suzy's eyesight. One might think that this is because mildly improper eyesight tends to bring about accidents only to a minimal degree, whereas heavy rainstorms have much bigger potential to bring about accidents. Or, while a mildly improper eyesight seldom brings about accidents, it's all too frequent that a heavy rainstorm brings about accidents. That is, a heavy rainstorm increases the probability of an accident much more than a mildly improper eyesight. One could then hold that a factor's degree of causal contribution is about how much that factor increases the probability of an outcome.[19] Hence, it seems plausible to hold that probabilistic theories of causation can account for degrees of causation.

To be sure, I don't mean that the accounts of causation above entail or even strongly suggest that there are degrees of causation. But these accounts don't inherently exclude a graded notion of causation either. Hence, I conclude that the first objection fails to raise a significant worry for the thesis that causation comes in degrees.

3. Objection 2: No Room for Thinking of More or Less of a Cause

[19] Cf. Kaiserman (2016) for further details on an account along these lines. I mean to suggest neither that this is the only way probabilistic theories of causation can account for degrees of causation, nor that the above is the only conception of probability or probabilistic theory of causation. See Fitelson and Hitchcock (2011) for a survey of several other proposals for analyzing degrees of causation in terms of probabilities. See Sprenger (2018) for a discussion on how to adjudicate between these competing proposals and his argument favoring one of them over the others. Moreover, the three kinds of theories of causation discussed above are not the only kinds of theories of causation. For instance, there are also primitivist theories of causation—i.e., those that suggest that causation is unanalyzable. Although it depends also on how much such a theory can tell us about causation despite its primitivism, it's possible that a primitivist theory can also account for degrees of causation. See Moore (2012, pp.447-8, 2013, p.126), and Mumford (2013) for discussions on this.

We now turn to the second objection. Suppose ten teenagers of varying strengths come together to push a boulder down the hill. None of them is strong enough to push the boulder on their own, and each of them is needed to get the job done. But if the inputs of each of these teenagers are necessary and none is sufficient for the outcome, then it is misleading to think of any of them as more/less strong or important cause of the outcome. After all, if any of them weren't there, the boulder couldn't have been pushed down. None of them on their own could push the boulder down either. Hence, the objector contends, if each factor is causally necessary and none is causally sufficient for the outcome, then there's no room for thinking about any of them as more/less of a cause of the outcome.[20]

An initial worry about this objection might be that it's based on a specific kind of case—one in which multiple people jointly cause an outcome. But the point in the objection could be further generalized since anytime we cause an outcome, there are multiple causal factors at play. For instance, lighting a match takes many other causal factors in addition to dragging the match across the striker—appropriate temperature, oxygen in the air, and so on. Here again, each of these factors is causally necessary—in the above sense—and none is causally sufficient for the outcome. However, one might still worry that the point isn't general enough—that *not* in all cases where there are multiple causal factors each factor is causally necessary and none is causally sufficient for the outcome. Take, for instance, overdetermination cases. By definition,

[20] Thanks to Robert Van Gulick for raising this objection. Zimmerman (1985) raises a similar concern and argues that we should avoid the talk of causing an outcome in varying extents. See Northcott (2005a) for a similar concern and his reply to it.

they're the kind of cases where there are multiple sufficient causes for an outcome. Yet, I'll ignore this complication, and take that the objection raises a general enough concern.

The objection again is that, concerning a set of causal factors, if all of them are causally necessary and none is sufficient for an outcome, there's no room for thinking about any of them as more/less of a cause of the outcome. Yet, it's unclear why this should be the case. Suppose we need ten gallons of water in a tank. It so happens that Suzy has seven gallons of water, and Timmy has three, and no one else has any water. Suzy and Timmy pour the water they have into the tank, and we have the ten gallons of water. Both Suzy's action and Timmy's action are causally necessary—in the above sense—and none is causally sufficient for the outcome. But it seems appropriate to think that Suzy is more of a cause of the outcome, or she causally contributed more to the outcome, than Timmy. One might further think that this is because Suzy increases the probability of the outcome more, or because she transfers more energy to the outcome. The point is that, regarding a set of causal factors, even if all of them are causally necessary and none is sufficient for an outcome, they could still differ in respects that are plausible ways of thinking about degrees of causation. Hence, I conclude that the second objection fails to show that causation doesn't come in degrees.

4. Objection 3: The Notion of Degrees of Causation Is an Illusion

Let's now turn to the third objection. Sartorio (2020) argues that the thesis that causation comes in degrees is false. Her argument relies on two criteria for measuring degrees of causation that she suggests. I'll now present these criteria first and then her argument. Consider the following pair of cases:

***Bullet**:* You are the only shooter aiming at a victim. You shoot and the victim dies.

***Bullet-with-Bird**:* When you shoot, a bird collides with the bullet and slows the bullet down a bit, reducing its momentum in such a way that it alone is no longer enough to kill the victim. But the bird's flying by also independently dislodges a loose boulder that is not large enough to crush the victim to death on its own. Although neither the bullet nor the boulder is enough in itself to cause the death, *together* they are sufficient to cause it.

On the assumption that there are degrees of causation, it seems natural to think that you make a less substantial contribution to victim's death in *Bullet-with-Bird* than you do in *Bullet*. We can account for this thought via the following criterion:

(Sufficiency Criterion) How much a cause contributes to an effect is a matter of how close it comes to providing a *sufficient* condition for an effect.

In *Bullet*, your bullet on its own is sufficient to kill the victim. Whereas in *Bullet-with-Bird*, your bullet doesn't come as close to being sufficient to kill the victim. Hence, you are less of a cause of the death in *Bullet-with-Bird* than you are in *Bullet*.

Consider also the following two cases:

***Difference**:* You are the only shooter aiming at a victim. You shoot and the victim dies.

***No Difference**:* Unbeknownst to you, you are one of many shooters who are aiming at a victim. It takes one bullet to kill the victim. You shoot, and so does each of the other shooters. All the bullets reach the victim simultaneously, and the victim dies.

It seems plausible to think that you made a more substantial contribution to the victim's death in *Difference* than in *No Difference*. In *Difference*, the victim's death very crucially depends on you, whereas in *No Difference* the victim would have died even if you weren't there. We can account for this thought via the following criterion:

(Necessity Criterion) How much a cause contributes to an effect is a matter of how close it comes to providing a *necessary* condition for an effect.

In *Difference*, your bullet is necessary to kill the victim—whether the victim lives or dies hinges on what you do. Whereas in *No Difference*, your bullet is one among many that are sufficient to kill the victim—it's far from being necessary for the victim's death. Hence, you are more of a cause of the victim's death in *Difference* than in *No Difference*.

The trouble begins, Sartorio argues, once we contrast *Bullet-with-Bird* with *No Difference*. Notice that your bullet is sufficient to kill the victim in *No Difference*, but not in *Bullet-with-Bird*. Notice also that your bullet is necessary for the victim's death in *Bullet-with-Bird*, but not in *No Difference*. It follows that according to **Sufficiency Criterion** you're more of a cause of the victim's death in *No Difference*, but according to **Necessity Criterion** this is false. And according to **Necessity Criterion** you're more of a cause of the victim's death in *Bullet-with-Bird*, but according to **Sufficiency Criterion** this is false. Pointing out this mismatch, Sartorio raises the question as to in which case you actually are more of a cause. In response, she says:

> Here I am at a loss about what to say. I feel like I just do not have any clear intuitions anymore. I find myself wanting to say: well, in a sense, [the] contribution is more significant in [one] case, and in another sense it is more significant in the [other] case. But this is unhelpful. (Sartorio 2020, p.352)

She concludes that the idea that causation comes in degrees is an illusion.[21]

However, this conclusion is unwarranted. It seems premature. We don't typically think that if two accounts of *x* sometimes give us conflicting results about *x*, (it's good reason to think that) *x* is an illusion. Take, for instance, two principles of morally right action—the utilitarian

[21] **Sufficiency Criterion** and **Necessity Criterion** might turn out to be a single criterion which might eliminate the mismatch in question. There's also some empirical support for the idea that the best way to account for the seemingly competing causal judgements based on necessity and sufficiency is via a unified account of degrees of causation (Icard, Kominsky, and Knobe 2017). However, this may not be enough to respond to Sartorio. As she grants it, there are still other plausible criteria for measuring degrees of causation, and she would contend that she could come up with other examples of the same kind of mismatch using them instead (cf. Sartorio 2020, p.351).

principle and the Categorical Imperative. Sometimes they give us conflicting results about whether an action is morally right. But hardly anyone concludes because of this that moral rightness is an illusion.

Indeed, more similarly to the subject at hand, theories of causation sometimes conflict in their verdicts about whether a given factor is a cause of an outcome or not. Take a simple counterfactual theory of causation according to which, again, *c* is cause of *e* just in case had *c* not occurred *e* would not have occurred. This theory can't properly account for overdetermination, since in overdetermination there are multiple sufficient causes.[22] Take any of the factors involved in an overdetermination case, it's false that had that factor not occurred the outcome wouldn't have occurred. Hence, we get the result that for any given factor involved in overdetermination, that factor isn't a cause of the outcome. But, according to one prominent productive theory of causation, causation is a matter of energy transference between a cause and an effect. Hence, if there's energy transference between a factor and the outcome in an overdetermination case, that factor counts as a cause of the outcome. Plausibly then each of those factors involved in that overdetermination case counts as a cause of the outcome (Moore 2009, p.105). For another well-known example, while productive theories typically can't count omissions as causes, counterfactual theories typically can (Paul and Hall 2013, pp.190-5). Consequently, some think that this counts against productive theories of causation while some think that omissions can't be causes. We can further multiply these sorts

[22] Cf. Moore (2009, p.354) and Dowe (2013, pp.115-6) for brief discussions on this. For ease of exposition, I use the simple counterfactual theory of causation. But even the more sophisticated counterfactual theories of causation can't properly account for overdetermination (Paul and Hall 2013, pp.149-51).

of examples. The point is that no one concludes based on these considerations that causation is an illusion.

To be fair, Sartorio's argument also includes explaining away "the appearances" in some cases where we need degrees of causation. Hence, as Sartorio (in personal communication) draws my attention to it, her argument is of the following kind: Here's a puzzle about *x*, and the best way out of it is explaining away the appearances about *x*. However, this still seems problematic to me for three reasons. First, as I'll show in section six below, the kind of cases where we need degrees of causation span much wider. Second, it seems premature to call for explaining away the appearances based on the given reason. It's ubiquitous in philosophy that competing theories on a given subject sometimes give us conflicting results. Compatibilist and libertarian theories sometimes give us conflicting results about whether someone is morally responsible. Internalist and externalist theories of epistemic justification sometimes give us conflicting results about whether a given belief is justified. The utilitarian principle and the Categorical Imperative sometimes conflict in their verdicts about whether a given action is morally right. But hardly anyone calls for explaining away moral responsibility, or epistemic justification, or moral rightness because of this. Third, the call for explaining away the appearances seems premature also because degrees of causation hasn't received much attention in the causation literature yet (Tadros 2018, p.408). Due partly to this lack of attention, Mumford says that he would hope that all *future* theories of causation take as a desideratum that degrees of causation must be accommodated (Mumford 2013, p.111, my emphasis). I conclude then that the third objection doesn't show that causation doesn't come in degrees.

5. It Is Too Costly to Deny That Causation Comes in Degrees

I presented the three main objections from the literature to the thesis that causation comes in degrees and argued that they fail. Now why should we think that this thesis is true? Relatedly, what is the cost of denying this thesis?

Let's begin by noting that our everyday thought and talk make extensive use of a graded notion of causation. We often wonder about degrees of 'causal potency,' of 'causal contribution,' of 'causal efficacy'; something being a 'main,' 'chief,' or 'principal' cause of an outcome; something being 'stronger/weaker,' 'more/less important,' cause of an outcome. For instance, one might wonder if it's the bad road conditions or poor driving skills that's more of a cause of traffic accidents. One might wonder whether it's the traffic or his driving skills that contributes more to his being often late to work. A high school student might believe that it's more because of her math teacher than any other of her teachers that she's graduating with honors. A young novelist might hold that Emily Brontë is the most important reason why he's a novelist.[23]

Note that these sorts of inquiries and beliefs also affect our attitudes and behaviors in very specific ways. Plausibly, if it's the bad road conditions that's more of a cause of accidents, we would prioritize investing into fixing the roads. If it's more or mainly due to your poor driving skills that you're often late to work, you might want to invest into improving your driving skills. If it it's more or mainly due to traffic that you're often late, you might want to

[23] Ordinary expressions of causal judgements—especially taken on their own—might be approached with caution for they might be indicative of that which is little metaphysical significance. However, it should be a virtue of a view if it's consistent with such expressions. Also, a growing body of psychology literature presents considerable empirical evidence that people's causal judgements themselves presuppose a graded notion of causation (cf., e.g., Lagnado et al., 2014, Gerstenberg et al., 2018, Langenhoff et al., forthcoming).

consider relocating instead. The high school student above presumably will respect her math teacher more than any other of her teachers, and might even plan on keeping in touch with the math teacher after graduation. The young novelist above might be willing to buy a well-preserved, first edition copy of Emily Brontë's classic novel *Wuthering Heights* even if the price is exorbitant. Note that these thoughts and attitudes seem intuitively very plausible. Note also that it's hard to see how we could properly make sense of these thoughts and attitudes if we deny that causation admits of degrees.

A graded notion of causation also undergirds many substantial scientific, legal, and philosophical theses. For instance, a historian might think that nationalism was more of a cause of the First World War than militarism. A physicist might think that gravity is more of a cause of a particle's acceleration than the presence of an electric field. A medical scientist or a physician might think that one medicine is more causally effective in curing a certain illness than another, or wonder how causally powerful a certain carcinogen is to kill patients.[24] Consider also the very live debate, among experts from various scientific fields, on what factor (e.g., which country, politician, or policy) is more of a cause of the Covid-19 pandemic death toll. These sorts of theses and inquiries seem perfectly sensible, and again they have wide ranging affects—from our personal attitude towards a politician to a country's position in international relations.

Moreover, how significant a causal role one plays in generating an outcome can be relevant to the degree of one's moral duty in reparations, or one's legal liability, regarding that

[24] Cf. Northcott (2005a, 2005b, 2013), Korb, Nyberg, and Hope (2011), Kaiserman (2016, 2018), Sprenger (2018) for further examples and more detailed discussion.

outcome (Miller 2001, Moore 2009, Kaiserman 2017). Whether it was the rain or a driver's degree of drunkenness that was the main cause of an accident can be relevant in personal or legal disputes. Of the two relevant factories, how much each of them contributed to the pollution of a river can be relevant for the question regarding reparations. It's again difficult to see how we can properly make sense of any of these thoughts and theses if we deny that causation comes in degrees. It seems then that a graded notion of causation is intuitively very plausible, and its denial is theoretically too costly.

Now, I don't deny that there might be competing explanations—ones that don't appeal to degrees of causation—for (at least some of) the thoughts, theses, and attitudes above. One might wonder, for instance, whether what undergirds the scientific theses above is a notion of degrees of explanatory importance rather than degrees of causation.[25] However, firstly, considering the long list of a wide range of thoughts, theses, and attitudes above, it would be highly surprising if it turned out that all the items in the list are accounted for without appealing to degrees of causation. Secondly, even if there are such competing explanations for all of them, considerations from theoretically unity would speak in favor of degrees of causation. This is because while the degrees of causation explanation provides a single principle to account for all of them, it's more likely than not that these competing explanations would be different in kind and not grounded in a single principle—considering, again, the range of items in the list. And thirdly, given that a graded notion of causation is intuitively plausible and all the main

[25] Thanks to an anonymous referee from *Synthese* for suggesting this potential explanation. (However, cf. e.g., Sprenger (2018) who observes a distinction between degrees of explanation and degrees of causation, and still contends that a graded notion of causation undergirds many scientific theses.) Thanks also to Carolina Sartorio for raising the concern that there might be other explanations for the thoughts, theses, and attitudes mentioned above.

objections against it fail, the burden of proof—to show that the items in the list don't presuppose graded causation—is on those who deny it.

6. Conclusion

I argued that the three main objections in the literature against the thesis that causation comes in degrees fail. I also showed that this thesis undergirds a wide-ranging portion of our everyday thought and talk, our small- or large-scale attitudes, and a variety of substantial scientific, legal, and philosophical theses. In short, we don't have good reasons to reject this thesis, and it's too costly to deny it. Hence, I conclude that causation comes in degrees.

Indeed, one might still wonder why some authors would oppose this thesis if its implicit or explicit use is so widespread as I suggest it. I suspect it's largely because this thesis hasn't received much attention in the causation literature yet. I think another reason why some might be suspicious about this thesis is that—as the foregoing discussions suggests—it's not always clear what's meant by degrees of causation to begin with. I hope that the discussion in this chapter also raises an awareness about this problem in the literature and remedies it to an extent.

Chapter 2: Moral Responsibility Is Not Proportionate to Causal Responsibility

Abstract:

It seems intuitive to think that if you contribute more to an outcome, you should be more morally responsible for it. Some philosophers think this is correct. They accept the thesis that *ceteris paribus* one's degree of moral responsibility for an outcome is proportionate to one's degree of causal contribution to that outcome. Yet, what the degree of causal contribution amounts to remains unclear in the literature. Hence, the underlying idea in this thesis remains equally unclear. In this chapter, I'll consider various plausible criteria for measuring degrees of causal contribution. After each of these criteria, I'll show that this thesis entails implausible results. I'll also show that there are other plausible theoretical options that can account for the kind of cases that motivate this thesis. I'll conclude that we should reject this thesis.

Suppose you and two of your friends want to plan a conference together. But your friends tell you that they won't do as much work. They'll help you but leave the work mostly to you. You agree and all three of you organize the conference together. It seems that you're *more* morally responsible—that is, more praiseworthy—for the outcome than they are. If asked for a reason why, it seems natural to reply that it's because you contributed more to the outcome.

Here's another case. Suzy and Billy are making a big birthday cake for their friend. Billy only puts a couple of cherries on top, while everything else is made by Suzy. It seems that Suzy is more praiseworthy for the outcome than Billy, and that this is because Suzy contributed more to the outcome.

It's typically accepted at least two factors—control and awareness—affect moral responsibility. Roughly, to be morally responsible for an action, one must perform that action

freely and be aware of various morally relevant factors concerning that action. Many philosophers hold, in addition, that what one causes also affects one's moral responsibility.[26] Yet the exact relationship between moral responsibility and causation is a further question. The two cases above suggest the intuitive idea that one's degree of moral responsibility for an outcome should be proportionate to one's degree of causal contribution to that outcome. Some philosophers hold, moreover, that this is correct.[27] They hold that:

> (**Proportionality**) *Ceteris paribus*, the more/less one causally contributes to an outcome, the more/less one is morally responsible for that outcome.[28]

Indeed, if we assume that causation, or causal contribution, doesn't come in degrees, **Proportionality** is false or uninteresting. Similarly, assuming that causation doesn't affect moral responsibility, **Proportionality** is false again. Even after denying these assumptions, however, **Proportionality** faces a major obstacle. This is because after granting that causation comes in degrees, we face the task of showing how to measure degrees of causation. So, one, without a further account of measuring degrees of causation, the underlying idea in **Proportionality** is not

---

[26] Cf. Moore (2011:496), Rosen (2011:405), Mumford (2013:109-10). While Moore (2011:496) says "roughly one-half of the philosophic community" accepts it, Mumford (2013:109) even says that the idea is "perhaps… too self-evident to require an argument at all." It should also be noted that the thesis that causation is necessary for moral responsibility is widely held (cf. e.g., Feinberg 1968:674, Kaiserman 2018:5). Although she rejects this thesis, Sartorio (2004:316) also notes that the thesis is widespread among philosophers.

[27] Cf., e.g., Braham and van Hees (2009:324), Moore (2009:71-2), Mumford (2013:109-10), Bernstein (2016:446), Bernstein (2017:167), Tiefensee (2019:242). After Bernstein (2017), I'll call this thesis **Proportionality**. It's also noteworthy that for some philosophers **Proportionality** is the default position *assuming that* causation affects moral responsibility or that causation comes in degrees (cf. Sartorio 2010:836-7, Alexander 2011:307-8, Alexander and Ferzan 2012:84, Sartorio 2015a:140). Moreover, it's sometimes argued that *legal responsibility* is proportionate to causal responsibility (Kaiserman:2017), but I won't take a position on legal responsibility.

[28] This *ceteris paribus* clause isn't always clearly expressed. Sometimes the context suggests this reading (e.g., Sartorio 2015a:140ff), and sometimes it's rather clear in the text (e.g., Moore 2009:71, Bernstein 2017:166). Petersson (2013) rejects **Proportionality** without the *ceteris paribus* clause.

clear enough. And two, once we have such an account, **Proportionality** might turn out to be *not* as plausible as it might initially seem.

But we face yet another obstacle. There are only several developed criteria for measuring degrees of causation in the literature. In my discussion below, I'll use these criteria. I'll also motivate and suggest various other criteria which should be a modest contribution to the growing literature on degrees of causation irrespective of their function in the context of **Proportionality**. In **§2** through **§4**, I'll consider six criteria in total and show that **Proportionality**, understood after each of these criteria, entails implausible results. In **§5**, I'll argue that we don't need **Proportionality** in order to account for the kind of cases that motivate this principle. Hence, I'll conclude, we should reject **Proportionality**.

Here's one caveat. The criteria that I'll consider are necessarily underdeveloped and have some obvious shortcomings. For instance, those that are inspired by productive theories of causation might have a hard time accounting for causation by omission. The ones inspired by counterfactual theories of causation might have a hard time accounting for causal overdetermination. The one inspired by probabilistic theories of causation might have a hard time accounting for event-token causation. However, causation by omission is a well-known problematic case for productive theories of causation. Likewise, overdetermination for counterfactual theories of causation, and event-token causation for probabilistic theories of causation. Surely, there are theoretical options around these problems. For instance, regarding overdetermination, one among the various ways of understanding it (Schaffer 2003), or completely rejecting it (Bunzl 1979; Tuomas 2018), might be among the options. But I won't

tackle these problems. I'll assume that solutions that can be employed by these theories of causation will apply also to the criteria that I consider.

Before we begin the discussion, however, let's briefly clarify again the notions of causal responsibility and moral responsibility in question.

1. Causal Responsibility and Moral Responsibility

Causal responsibility occurs when one event causes another.[29] Throwing a rock might be causally responsible for breaking a window, rain for a flood, fire for the destruction of a forest. Causal responsibility merely highlights the causal involvement of one thing with another. One could be causally responsible for an outcome without deserving any blame or praise for that outcome. Suppose you push a door open to walk into a building. Unbeknownst to you, someone is standing idly on the other side of the door, and you knock him down. Plausibly you're causally responsible for his falling down, but not morally responsible—i.e., not blameworthy—for it.

"To be morally responsible" might mean two things. One, roughly, is to have a moral duty. For instance, parents are morally responsible for taking care of their children, teachers are responsible for imparting knowledge to their students, and so on. And two, it might mean being apt for moral praise or blame. The kind of moral responsibility that's relevant to **Proportionality** is the latter one. Hence, hereafter by "moral responsibility," I'll refer to the latter one.

[29] The question of causal relata is controversial. Events, facts, states of affairs, and features are among the possible candidates. The standard view holds that they're events (Paul and Hall 2013, Schaffer 2016). I don't take a position on this. However, for ease of exposition, I'll treat causation as a relation between events.

One thing to note about our judgements of moral responsibility is that they often involve comparisons—not just within a case that involves multiple agents, but also across distinct cases. For instance, we hold that someone who told a trivial lie, or has an outstanding parking ticket, is less blameworthy than a murderer. Similarly, we hold that two people who killed their respective victims are equally blameworthy. These judgements seem to be justified by a widely accepted, more general idea—that like cases should be treated alike. For instance, if Suzy commits a murder, and Billy commits a murder, *ceteris paribus* Suzy and Billy are morally responsible to the same degree. And whether the victim is male or female, young or adult or old, healthy or sick, rich or poor, etc. seems irrelevant to the degree to which Suzy and Billy are morally responsible. Indeed, there's a lot more to be said here. But this should suffice for my purposes. In my discussion below, I'll employ comparisons across distinct cases, but I'll stay away from the kind of cases that might be controversial. The point here is largely to draw attention to our judgements of moral responsibility across distinct cases, and show that these judgements are not without any basis.

2. Productive Theories of Causation

We can now begin considering various criteria for measuring degrees of causation and test **Proportionality** by employing them. In the discussion below, various reminders from the previous chapter will be necessary. But I'd rather not use phrases like "as mentioned in such and such chapter" more times than is necessary. So, I'll write as if they occur for the first time.

Now imagine that three people carrying a heavy marble top, rectangular dining table from the truck all the way into the kitchen. Two of them, who are regular people, each hold two

separate corners on one side, and one of them, who is unusually large and strong, holds both corners on the opposite side. It seems natural to think that the strong one causally contributed more to the outcome than the other two people. A plausible way to explain this thought would be by appealing to the amount of input from the strong person. After all, she did on her own what two other people did together. She spent much more effort, or put more force, than the other two in carrying the table. A criterion for measuring degrees of causation that's inspired by a productive theory of causation can account for this thought.

Productive theories of causation hold that causation is a matter of producing another event. According to one prominent example of such an account, *c* causes *e* if there's a transfer of conserved quantity (for instance, force, or energy) from *c* to *e* (Dowe 2000). Suppose you hit a glass and it falls off the table. You brought about this outcome in virtue of transferring energy to the glass. In the dining table case above, one could hold that the strong person causally contributed more to the outcome because she transferred more energy to the outcome. Hence, a criterion for measuring degrees of causation that naturally follows is this:

> **(CC1)** The amount of causal contribution to an outcome is the amount of energy transferred to the outcome.[30]

Now let's consider cases regarding moral responsibility. Suppose it takes 10 units of energy to kill Victim1, and 20 units of energy to kill Victim2. Billy intentionally kills Victim1 by

[30] Moore (2012), Mumford (2013), and Bernstein (2017) consider accounts of causal contribution similar to this. However, notice that one might transfer more energy than needed for an outcome. Suppose it takes 10 units of energy to kill a victim, and one bullet amounts to 10 units of energy, but Suzy fires two bullets. Does that mean Suzy's causal contribution to the outcome is 20 units of energy? Not necessarily. One might hold a sort of threshold view according to which the excess amount of energy transfer doesn't count as one's causal contribution to the outcome. 10 units of energy transfer brings about the death, and the excess 10 units doesn't contribute to the death, since one doesn't make the victim more dead by transferring more energy than needed for the death.

transferring 10 units of energy, and Suzy intentionally kills Victim2 by transferring 20 units of energy. Suzy's causal contribution to Victim2's death is twice the amount of Billy's causal contribution to Victim1's death. If **CC1** is correct, then **Proportionality** tells us that that Suzy is more morally responsible than Billy. But this seems just false. They both intentionally committed a murder, and seem to be equally blameworthy.[31] Thinking otherwise commits one to hold, for instance, that in cases where murderers use just brute force to kill their victims (say, by smothering), killing children and the disabled typically makes one much less blameworthy than killing healthy adults. This is because it would typically take less energy transfer to kill someone in the former set. But it is clearly wrong that killing a child makes one less blameworthy than killing an adult. We can conclude then that **Proportionality** understood after **CC1** is false.

However, one might object that the initial thought above regarding the dining table case can be accounted for in a slightly different way. Suppose it takes 30 units of energy in total to carry that table. Suppose also that the two regular people each transfer five units of energy to the outcome, and the strong person transfers 20 units of energy. The ratio of the strong person's energy transfer to the energy transfer needed for the outcome is 20/30. The ratio of the each of the other two people's energy transfer to the energy transfer needed for the outcome is 5/30. If we take causal contribution along the lines of these ratios, we still get the

[31] As required by the *ceteris paribus* clause in **Proportionality**, I assume here that Suzy and Billy are on a par not only concerning (the strength of) their intentions but also concerning other factors that might be relevant to moral responsibility. That is, I assume that they're on a par concerning their epistemic statuses, that they're not under duress, that they're not manipulated, and so on. The only relevant difference between them is the degree of their causal contributions. Below, I will refrain from repeatedly elaborating on these details in the description of my counterexamples. The counterexamples are designed to satisfy the *ceteris paribus* clause.

conclusion that the strong person causally contributed more to the outcome than the other two. And we still get this result because of the amount of effort from the strong person just as the initial thought above suggested. Moreover, **Proportionality** understood after this way of measuring degrees of causation will give us the correct result in Billy and Suzy killing Victim1 and Victim2 respectively. Billy's causal contribution will be 10/10, which is one, and Suzy's causal contribution will be 20/20, which is also one. Since the amount of their causal contributions will be the same, **Proportionality** tells us that they're both equally morally responsible. And this is the conclusion that we wanted. Hence, the following could be a better criterion for measuring degrees of causation:

> **(CC2)** The amount of causal contribution to an outcome is the ratio of the amount of transferred energy to the amount of energy needed for the outcome.

But consider now the following case. Suppose Victim3 is standing at the edge of a cliff. It will take only three units of energy transfer for Victim3 to fall and die. Suzy and Billy, who have been planning to kill Victim3, each give a slight push to Victim3 with their fingertips. Suzy's push was relatively gentler and happened to transfer one unit of energy, while Billy's push transferred two units of energy. **CC2** tells us that Suzy's causal contribution to Victim3's death is 1/3 whereas Billy's causal contribution is 2/3—Billy's is twice as much as Suzy's. And if **CC2** is correct, then **Proportionality** tells us that Billy is much more morally responsible than Suzy. But it seems wrong that Billy deserves a considerably higher degree of blame. They seem equally blameworthy. Hence, **Proportionality** understood after **CC2** is false.

3. Dependent Theories of Causation

Imagine Billy and Suzy working in an apple orchard. Today, they're asked to pick up 1000 apples that are ripe, firm and with no indents or discoloration. At the end of the day, Billy counts 400 apples in his basket, and Suzy 600. They bring the apples to the office and call it a day. It is natural to think that Suzy causally contributed more to the outcome than Billy. Moreover, it seems plausible to think that Suzy contributed more even if Suzy and Billy spent equal amount effort. They both worked nonstop, they are equally tired in the end, they spent equal amount of energy and so on. After all, Suzy might just have worked smart rather than hard, and hence got more work done. If this is right, then **CC1** and **CC2** above can't properly account for the thought that in this case Suzy causally contributed more than Billy. The difference in the degrees of their causal contribution here isn't about their inputs or efforts, but what their inputs effectively translate to in terms of the outcome. Put differently, the difference seems to be about the portion of the outcome that depends on each of their conducts. A criterion for measuring degrees of causation that's inspired by a dependent, or counterfactual, theory of causation can account for this thought.

According to counterfactual theories, causation is a matter of counterfactual dependence between wholly distinct events (Hall 2004). On the simple counterfactual view *c* is a cause of *e* just in case *e* counterfactually depends on *c*. That is, *c* causes *e* just in case had *c* not occurred, *e* would not have occurred. The dependence relation in question doesn't initially seem to leave room for gradation: either the counterfactual dependence holds or not. Hence,

one might be inclined to hold that the simple counterfactual cannot account for degrees of causal contribution.[32] But this is mistaken. Consider the following criterion:

> **(CC3)**: The amount of causal contribution is the portion of the outcome that counterfactually depends on an agent's conduct.

Take again Billy and Suzy picking apples. The portion of the outcome that counterfactually depends on Billy's conduct is 400 hundred apples. If not for his conduct, their daily objective would be 400 apples short. Similarly, the portion of the outcome that counterfactually depends on Suzy's conduct is 600 hundred apples. **CC3** then tells us that Suzy's causal contribution to their daily objective is more than Billy's causal contribution. And this is so because the bigger portion of the work depends on Suzy's conduct and regardless of how much effort each of them spent.

Now consider the following case. Suzy and Billy are great friends. They are equally talented, they know each other's ways, and spontaneously work well as a team. They decide that they'll start a defamation campaign against Victim3. They each individually start making up material that will eventually be used in their defamation video. Since they're equally talented and typically complement each other well, they're not worried that their separate materials won't add up to a coherent video. They each put equal amount of thought and effort into their work and come up with equally useful materials. But not all their materials can make the final cut. They make an arbitrary choice and a couple more of Billy's material than Suzy's material end up being in the video. They finalize the video and upload it on internet turning Victim3's life

[32] See Moore (2009), Mumford (2013). Moore seems to have changed his mind on this (cf. Moore 2012).

upside down. Now, since Suzy has less material to be used in the final video than Billy, the portion of the outcome that counterfactually depends on Suzy's conduct is smaller than that of Billy's.[33] **CC3** then tells us that Suzy's causal contribution to the outcome is less than that of Billy's. Hence, **Proportionality** understood after **CC3** tells us that Suzy is less blameworthy for turning Victim3's life upside down than Billy. But this seems false. We wouldn't have, for instance, thrown Suzy into jail for five years and Billy for seven. Or, if the proper reaction to what they did is to rehabilitate them, we wouldn't have Suzy kept in rehabilitation for less time, or rehabilitate her any differently, than Billy. Hence, **Proportionality** understood after **CC3** is false.

However, the initial thought above regarding the apple picking case can be account for in another way. Suzy's picking up 600 apples, as opposed to Billy's 400 apples, might be an indication of Suzy's being a more talented apple picker. Being more talented implies being more irreplaceable at what one does which in turn could imply more causal contribution. Consider an artwork like Taj Mahal. The artist that designs the artwork and a manager in charge of the food supplies for the workers both causally contribute to the artwork. But it seems that the artist contributes more than the manager. The artist has a unique set of skills compared to the manager. There are many people who can do what the manager does, but not as many people

[33] If we take the outcome as the final video, it's clear that the portion of the outcome that depends on Suzy's conduct is smaller. But if we take the outcome as the harm to Victim3, we need further details. (Thanks to Ben Bradley for drawing my attention to this.) We could stipulate that Victim3 would be somewhat better off had it not been for the relatively smaller portion of Suzy's contribution to the final video. But Victim3 would be much more better off had it not been for the relatively bigger portion of Billy's contribution to the final video. Hence, we could again plausibly think that the portion of the outcome, i.e. the harm to Victim3, that depends on Suzy's conduct is smaller than that of Billy's. (Thanks to Neil Feit for the helpful discussion on this.) Now, it seems to me that Suzy and Billy can be blameworthy for merely producing that video even if no harm consequently befalls Victim3. Moreover, regardless of which outcome we have in mind, they both seem equally blameworthy. Hence, on either interpretation of the case, I have a counterexample to **Proportionality** understood after **CC3**.

who can do what the artist does. The artist is much more irreplaceable than the manager. That is, had the manager not existed, her place would relatively easily have been filled by another. But had the artist not existed, her place would not have so easily been filled by another. Hence, in some sense of the word, the artist is more necessary for the artwork than the manager. Put differently, the artwork depends more crucially on the artist than the manager. It's plausible to think that this sort of being more necessary, or more crucial, means more causal contribution. It would explain why we think that the artist causally contributes more to the outcome. It would also give us the intuitively correct result that Suzy causally contributes more than Billy in the apple picking case, and explain why this is the case. Here then is another plausible criterion for measuring degrees of causation:

> **(CC4)**: The amount of causal contribution is the degree to which the outcome crucially depends on an agent's conduct.

The idea of crucialness here, again, is to be understood in terms of irreplaceability—how irreplaceable the agent is at what he does.[34] The more irreplaceable the agent is at what he does, the more he causally contributes to the outcome. Put differently, the more talented the agent is at what he does, the more he causally contributes to the outcome.[35] Notice that being more irreplaceable, or more talented, at one's job doesn't always correspond to getting a bigger portion of the outcome done. For one thing, one could be the most talented, or the most irreplaceable, in getting a smaller portion of a task done. For another, two people who are

---

[34] Cf. Lagnado, Gerstenberg and Zultan (2014:1055-6), Tadros (2018:411-3), for various other ways the basic idea here might be understood.

[35] Being highly talented doesn't always imply being highly irreplaceable. For instance, most people are highly talented at various simple tasks like flipping a switch. Hence, what I have in mind here is not mere high talent but *being more talented than many or most* at what one does which plausibly always implies being highly irreplaceable. Thanks to Ben Bradley for raising a concern that helped clarify myself here.

equally talented, or equally irreplaceable, in what they do might end up getting varying portions of a task done. Recall, for instance, that Suzy and Billy in the defamation campaign case above are equally talented but end up contributing to the outcome in varying portions. For this reason, notice also that **Proportionality** understood after **CC4** gives us the correct result in Billy and Suzy's defamation campaign case. Billy and Suzy, in that case, are equally talented—neither of them is more/less replaceable at what they do. **CC4** then tells us that Billy and Suzy equally causally contributed to the outcome. Hence, **Proportionality** tells us that Billy and Suzy are equally morally responsible for the outcome which is the correct result. **CC4** then is an independently motivated, distinct criterion for measuring degrees of causation that also might fare better than **CC3**.

However, consider now the following case. Suppose a mob leader wants two people, Victim5 and Victim6, kidnapped. Victim5 is a powerful and well protected person, whereas Victim6 is a regular person. The mob leader has many goons but only Suzy is sophisticated enough to plan Victim5's kidnapping. All other goons are equally sophisticated. The mob leader assigns Billy to plan Victim6's kidnapping. Suzy and Billy make their plans. Each plan requires, in addition to the planner, three equally well-skilled goons to execute the plan. The plans are executed and Victim5 and Victim6 are kidnapped. Suzy is much more irreplaceable at what she does than Billy is at what he does. **CC4** then tells us that Suzy's causal contribution to Victim5's kidnapping is much more than Billy's causal contribution to Victim6's kidnapping. And **Proportionality** understood after **CC4** gives us the result that Suzy is more morally responsible than Billy. But this seems unacceptable. They each planned and actively participated in kidnapping a person. It's hard to see why one should be any less/more morally responsible than

the other. If one holds that Suzy *is* more morally responsible, then one is committed to the following: kidnappers of weak and unprotected people are typically *less* morally responsible than kidnappers of powerful and protected people. This is because kidnappers of unprotected people would typically be people who are less irreplaceable at what they do then kidnappers of protected people. But this result is implausible. Hence, **Proportionality** understood after **CC4** is false.

I will now consider one last criterion based on counterfactual accounts of causation. One factor's causally contributing more/less to an outcome might also be thought of in terms of that factor's *influencing* the outcome more/less. Of course, a lot depends on how this notion of influence will be fleshed out. And no doubt more input to the outcome, or having bigger portion of the outcome done, or being more irreplaceable at what one does are some of the ways one might want to flesh out the notion of influence here. However, another way of thinking about influence might be along the lines of Lewis' (2004) account of causation as influence according to which more influence means more causal power. And influence tracks a pattern of dependence of, roughly, *the time and manner of the occurrence of a cause* upon *the time and manner of the occurrence of an effect*. Let's illustrate the pattern of dependence in question. Suzy throws a rock and shatters the bottle. Billy throws a rock which arrives immediately after Suzy's throw shatters the bottle. Lewis says that

> "altering Suzy's throw while holding Billy's fixed would make a lot of difference to the shattering, whereas altering Billy's throw while holding Suzy's fixed would not. Take an alteration in which Suzy's rock is heavier, or she throws a little sooner, or she aims at the neck of the bottle instead of the side. The shattering changes correspondingly. Make just the same alterations to Billy's preempted throw, and the shattering is (near enough) unchanged." (Lewis 2004:92)

Since the time and manner of the shattering is much more sensitive to the time and manner of Suzy's throw than it is to Billy's throw, Suzy's throw has much more influence. And more influence means more causal power. Accordingly, Suzy's throw causally contributes much more to the shattering than Billy's throw. Here then is another criterion for measuring degrees of causation:

> **(CC5)**: The degree of causal contribution is the degree of how sensitive the time and manner of the outcome is to an agent's conduct.

Consider now the following case.[36] Suppose Suzy and Billy, each pressing a button on a mechanism, will kill Victim7. Both buttons need to be pressed for Victim7 to die. The mechanism is such that the time and manner of Suzy's button being pressed makes no difference to the time and manner of Victim7's death if Billy's button is pressed. The time and manner of Billy's button being pressed makes significant difference to the time and manner of Victim7's death if Suzy's button is pressed. Suzy and Billy know all these facts and are indifferent to the potential variations in the time and manner of Victim7's death—whether his head will explode, or he'll be shot by a bullet, whether the mechanism will kick in in ten seconds or in a minute, and so on. They both push their buttons and Victim7 dies.

Among the criteria that are considered so far, only **CC5** clearly accounts for a difference in degrees of causal contribution in this case—if there's such a difference. We can safely assume that Suzy's button and Billy's button take the same amount of energy transfer. We can also safely assume that Suzy and Billy are equally talented at what they do—pushing their respective buttons. And, as I noted above, it's not clear how the talk of portions of an outcome

[36] I adapt this case from Tadros (2018).

should apply in this case for it's not clear how we could think of death in portions. **CC5**, however, clearly tells us that Billy's causal contribution to Victim7's death is much more than Suzy's causal contribution to Victim7's death for Billy has much more Lewis-style influence on Victim7's death. **CC5** then seems to be a distinct account of causal contribution. And **Proportionality** understood after **CC5** gives us the result that Billy is much more morally responsible than Suzy. But this seems implausible. Billy's having more Lewis-style influence on Victim7's death than Suzy doesn't seem to make Billy more morally responsible. After all, they were fine with the variations of the outcome that depended on the time and manner of pushing their respective buttons. Hence, **Proportionality** understood after **CC5** is false.

4. Probabilistic Theories of Causation

Consider the following scenario. Suzy had her six-monthly eye exam appointment yesterday, but missed it. She barely feels a problem with her glasses now, but they seem to work fine overall. And since her prescription didn't change in the last two years, she's not too worried about missing the appointment. She needs to drive thirty minutes to work but there's also a heavy rainstorm outside. She's a good driver, and she has driven in rainstorms before. Unfortunately, however, this time she gets into an accident on her way to work. It seems that the heavy rainstorm causally contributed more to the accident than Suzy's eyesight. Here's a plausible way to think why this is the case. All background conditions fixed, a heavy rainstorm alone brings things closer to an accident than a barely improper eyesight alone does. Granted, an improper eyesight tends bring about accidents. But a barely improper eyesight could tend to bring about accidents only to a minimal degree. Whereas heavy rainstorms have much bigger potential to bring about accidents. Plausibly while a barely improper eyesight seldom brings

about accidents, it's all too frequent that a heavy rainstorm brings about accidents. An account of causal contribution that's inspired by a probabilistic theory of causation can account for this thought.

According to probabilistic theories, causation is a matter of raising the probability of the occurrence of an event (Eells 1991). If *c* is a cause of *e*, then *c* raises the probability of the occurrence of *e*. A criterion for measuring degrees of causation that follows is this:

> **(CC6)**: The amount of causal contribution is the amount of increased probability, due to a given factor, of the occurrence of an outcome.[37]

In Suzy's case above, keeping all other background conditions fixed, the rainstorm alone made the crash more likely than her barely improper eyesight alone did. Accordingly, **CC6** tells us that the rainstorm causally contributed more to the crash than Suzy's eyesight.

Notice that increasing the probability of an outcome more doesn't necessarily correspond to putting more effort into that outcome. One might work smart rather than hard and increase the probability of the outcome more by spending less effort. Neither a factor's increasing the probability of an outcome more means that a bigger portion of the outcome depends on that factor. In Suzy and Billy's defamation campaign case above, both their efforts increased the likelihood of the final video to an equal degree. Suzy had equally good material prepared that made the final video equally likely to be produced. Neither is it that increasing

[37] See, Kaiserman (2016) for a more complex and developed version of this idea. Proponents of probabilistic theories of causation typically have in mind objective, as opposed to subjective, probability understood in terms of either propensity or frequency (cf. Hitchcock 2018). Accordingly, I take the probability in **CC6** to be objective probability. But I don't take position on whether probability is best accounted for in terms of frequency or propensity. Notice that I described Suzy's car accident case above both in terms of propensity and frequency.

the probability of an outcome more corresponds to being more irreplaceable at what one does. Suppose a traditional cobbler and an automated machine that produces shoes much faster are put to work for producing 1000 shoes. Suppose also that it's a time where there are few traditional cobblers but a lot of easy to obtain automated machines. The cobbler is much more irreplaceable at what he does, but the machine alone makes the outcome more likely than the cobbler does. Finally, increasing the probability of an outcome more is different than influencing the outcome more. We can safely assume that in Lewis's rock throwing case, Suzy and Billy are equally good at throwing rocks and that the rock each one throws is on a par with the other. Keeping all background conditions fixed, Suzy's throw and Billy's throw each alone will make the bottle's shattering equally likely while they still have the difference in their respective Lewis-style influences on the outcome. **CC6** then seems to be an independently motivated and distinct account of causal contribution.

Now, consider the following case. Billy is planning to kill two people by poisoning them. The probability that posion1 kills someone is .65. The probability that posion2 kills someone is .95. Billy adds poison1 to Victim8's drink, and posion2 to Victim9's drink. Both victims die soon after. **CC6** tells us that Billy's causal contribution to Victim9's death is more than Billy's causal contribution to Victim8's death. And **Proportionality** understood after **CC6** tells us that Billy is more morally responsible for killing Victim9 than for killing Victim8. This result seems implausible. It implies that murderers who use less potent poisons to kill their victims are less morally responsible than those who use more potent poisons. Also, consider for instance the fact that typical perpetrators' actions would make the kidnapping of well-protected kids less likely than the kidnapping of under-protected kids. We'd get the result that typical perpetrators

are less morally responsible for kidnapping well-protected kids than for kidnapping under-protected kids. This seems implausible. Hence, **Proportionality** understood after **CC6** is false.

5. On the Cases That Seem to Support Proportionality

None of the six criteria above for measuring degrees of causation vindicates **Proportionality**. If any of them is true, **Proportionality** must be false. This is considerable evidence to reject, or at least be suspicious of, **Proportionality**. However, one might worry that if we reject **Proportionality**, we might not properly account for cases that motivate **Proportionality**. Consider again the case where Suzy and Billy are making a big birthday cake for their friend. Billy only puts a couple of cherries on top, and everything else is made by Suzy. Suzy seems more praiseworthy, and it seems this is because Suzy contributed more to the outcome than Billy.

We seem to have two options. One, resist the argument above, and hope that **Proportionality** will come out true under another criterion for measuring degrees of causation. Two, look for other plausible ways of accounting for the cases in question. Let's firstly talk about the first option. Note that a cogent argument in favor of **Proportionality** won't only need to show that **Proportionality** is true under another criterion—*if* there is such a criterion. It will also need to show why all the criteria above are false. This would be a task of considerable weight. The prospects of success for such a task seems dim since I employed the already existing criteria and suggested new ones. Moreover, if we can account for the cases in question without **Proportionality**—to which I'll turn shortly—one might even worry that the hope for the first option is misplaced. Hence, the first option doesn't seem attractive.

Let's talk about the second option. There might be various ways of accounting for the cases in question. But what matters for my purposes is whether we can account for them after denying that causal responsibility is relevant for moral responsibility since I eventually want to argue that causal responsibility is irrelevant for moral responsibility. In the remainder of this section, I'll suggest that we can.

To begin with, we might wonder whether we should take our intuitions about these cases at face value. Recall that **Proportionality** includes a *ceteris paribus* clause. We might then wonder whether in the cases that motivate **Proportionality** everything else *is* equal. Recall, for instance, that the epistemic status of agents is among the factors that are relevant for moral responsibility. To illustrate again, suppose Suzy and Billy turn on a switch together. While Suzy knows that turning on the switch will wrongfully electrocute the person in the next room, Billy justifiably believes that it will only turn on the lights in this room. It seems that while Suzy is blameworthy for electrocuting the person in the next room, Billy isn't. Moreover, this difference seems sufficiently explained by the difference in their epistemic statuses.

So now we might wonder whether in the birthday cake case above Suzy and Billy are on a par regarding their epistemic statuses. Suppose the case is as follows concerning their epistemic statuses. Suzy believes that the birthday cake is for Jane, that Suzy and Jane are very close friends, that Suzy is good at making cakes, and that the birthday cake will make Jane very happy. So, Suzy (correctly) believes that she should make that birthday cake. And Billy believes that the birthday cake is for Jane, that Billy and Jane are merely acquaintances, and that Billy isn't good at making cakes. So, Billy (correctly) believes that it's nice of him if he helps Suzy with the cake but not wrong if he doesn't. Intuitively, none of these details change how

praiseworthy they are—Suzy still seems more praiseworthy than Billy. It would certainly still be appropriate if Jane feels more grateful to Suzy than to Billy.

Notice, however, that the case now violates the *ceteris paribus* clause, and hence doesn't support **Proportionality**. It's also now unclear how exactly the degree of their causal contributions should affect the degree of their praiseworthiness. **Proportionality** is no help with it. But it seems perfectly appropriate to explain why Suzy is more praiseworthy than Billy by appealing to the differences in their epistemic statuses and the further differences in their mental states that these epistemic differences indicate. While Suzy seems to 'care' strongly about making the cake and 'motivated' by the idea of making Jane happy, Billy is almost indifferent to making the cake and not as motivated by the idea of making Jane happy.

One might rightly wonder what happens if Suzy and Billy *are* on a par regarding their epistemic statuses and all other potentially relevant metal states, but differ regarding only the degree of their causal contributions. Suppose now, as Billy correctly believes, the birthday cake is for Jane, Billy and Jane are very close friends, Billy is good at making cakes, and the birthday cake will make Jane very happy. So, Billy and Suzy (correctly) believe that they should make that birthday cake together. It seems now that both Suzy and Billy 'care' equally about making the cake and are equally 'motivated' by the idea of making Jane happy. Still, all that Billy ends up contributing to the cake are a couple of cherries on top. Is Suzy now more praiseworthy than Billy? I'm not sure that she is.

My sense is that we either don't have a consistent story here or that they are both equally praiseworthy. Suppose despite all that's been said of him, Billy just sits there doing

nothing, shows up at the end, and adds the cherries on top. But now we don't have a consistent story because now it makes little sense to characterize him as caring equally about making the cake or equally motivated by the idea of making Jane happy. Suppose, on the other hand, Billy's contribution 'somehow' ends up being those couple of cherries on top. In that case, I'd want to hear more about that 'somehow.' To avoid having an inconsistent story like the one above, we might think something like this. An evil demon keeps randomly undoing Billy's contributions. If Billy stops working, the demon undoes Suzy's contribution too. Suzy and Billy realize this, and both keep contributing to the cake. Right at the end when Billy is about to add those cherries on top, the evil demon stops her devilry. So, Billy's contribution to the cake ends up being those cherries on top. And Billy and Suzy are on a par regarding all other potentially relevant factors. But now it seems hard to think of Suzy as more praiseworthy than Billy. They seem equally praiseworthy. And this is perfectly well explained by an appeal to their epistemic statuses and other relevant metal states.

To sum it up, when Suzy and Billy are *not* on a par regarding their epistemic statuses and other potentially relevant mental states, they seem differently praiseworthy. But, one, this doesn't support **Proportionality** because it violates the *ceteris paribus* clause in **Proportionality**. And two, the difference in the degree of their praiseworthiness can sufficiently be explained without any appeal to the degree of their causal responsibility. When Suzy and Billy *are* on a par regarding their epistemic statuses and other potentially relevant mental states, they seem to be equally praiseworthy. This further confirms the other counterexamples to **Proportionality** presented in previous chapters.

In chapters 4 and 5, I will say more on explaining moral responsibility without appealing to causal responsibility. But the discussion so far should suffice to show that plausibly we don't need **Proportionality** in order to account for the kind of cases that seem to motivate **Proportionality**. If this is correct, then **Proportionality** isn't as indispensable and attractive as it might initially seem. Considering also that **Proportionality** comes out false under all six criteria for measuring degrees of causation above, we have good reasons to reject **Proportionality**.

6. Conclusion

Proportionality suggests that there's a straightforward relationship between degrees of causal responsibility and degrees of moral responsibility. The more/less you causally contribute the more/less morally responsible you are. In order to test whether this relationship obtains, I employed various plausible criteria for measuring degrees of causation—those that are suggested in the literature and some that I suggested above. I argued that such straightforward relationship between causal responsibility and moral responsibility doesn't obtain under any of these criteria. I don't assume that I exhausted all the possible ways of measuring degrees of causation. But if none of these criteria vindicates **Proportionality**, this is considerable reason to reject **Proportionality**. I also argued that plausibly we don't need Proportionality in order to account for the kind of cases that motivate **Proportionality**. All this should amount to good evidence that **Proportionality** is false.

Chapter 3: Against Resultant Moral Luck

Abstract:

Does one's causal responsibility increase the degree of one's moral responsibility? The proponents of resultant moral luck hold that it does. Until quite recently, the causation literature has almost exclusively been interested in the binary question of whether one factor is a cause of an outcome. Naturally, the debate over resultant moral luck also revolved around this binary question. However, we've seen an increased interest in the question of *degrees* of causation in recent years. And some philosophers have already explored various implications of a graded notion of causation on resultant moral luck. In this chapter, I'll do the same. But the implications that I'll draw attention to are bad news for resultant moral luck. I'll show that resultant moral luck entails some implausible results that leave resultant moral luck more indefensible than it was previously thought be. I'll also show that what's typically taken to be the positive argument in favor of resultant moral luck fails. I'll conclude that we should reject resultant moral luck.

Does one's causal responsibility increase the degree of one's moral responsibility? The proponents of resultant moral luck (**RML**) hold that it does. Consider two reckless drivers, Lucky and Unlucky. Lucky gets into her car, drives home safe and sound. However, Unlucky isn't as lucky. On her way home, a kid jumps into the road, and she kills the kid. The proponents of **RML** hold that both Lucky and Unlucky are blameworthy, but Unlucky is much more blameworthy than Lucky. This is despite the fact that what they end up being causally responsible for is not up to them. Whether someone jumps into the road or not is out of their control.[38]

[38] Resultant moral luck is one among the various kinds of moral luck. Cf. Nagel (1979) and Williams (1981) for the classic discussions on moral luck. The term "resultant luck" is due to Zimmerman (1987). For many, resultant moral luck is the most obviously problematic kind of moral luck (Sartorio 2012). Cf. Nelkin (2019) for the recent state of the debate over moral luck.

The often-used positive argument in favor of **RML** is that our ordinary moral assessments support **RML**.[39] For instance, it's argued, we ordinarily find drunk drivers who end up killing someone much more morally blameworthy than those who don't kill anyone. Similarly, we find murderers significantly more blameworthy than those who merely attempted murder—even if the lack of success in the murder attempt is due to the fact, say, that the victim tripped and fell right before the bullet arrived.

As I see it, **RML** is the idea that *all else equal* causal responsibility increases one's moral responsibility. This characterization perfectly captures the commitment of the defenders of **RML** in the case of Lucky and Unlucky above and in other paradigm examples of **RML** cases. Some philosophers want to emphasize the role of "luck" in characterizing **RML**. Consider this: **RML** occurs when someone is more/less morally responsible for an action even if what she ends up bringing about is partly due to luck. But I find such characterizations unhelpful because they unnecessarily complicate the debate. In thinking about the case of Lucky and Unlucky, we want to know how morally responsible each of them is and the only difference between the two is that one is causally responsible for something that the other isn't. And as the "even if" part in this second characterization of **RML** indicate, whether this difference is due to luck is irrelevant. Nothing too significant in my discussion depends on one characterization or the other. I can run my argument also by employing the second one. But as I mentioned, the first one keeps the

[39] Compare: "to reject the judgments and practices that seem unavoidably to lead to [moral luck] would require a radical […] revision of ordinary moral evaluation" Wolf (2001:5). Cf. also Moore (2009:29-30), Mumford (2013:109-10), Kneer and Machery (2019), and Nelkin (2019). It might turn out that people's ordinary assessments don't support **RML**. In fact, Kneer and Machery (2019) study shows that they don't. I'll argue that these purported ordinary assessments lead to contradictory results and hence should be given up. So, if I'm right, even if it turns out that people's ordinary assessments are as the proponents of **RML** claim they are, this still can't support **RML**.

discussion much clearer. Moreover, if the first characterization is false, the second one is also false. Hence, in my discussion below, I'll focus on the first one.[40]

Until quite recently, the causation literature has almost exclusively been interested in the binary question of whether a given factor is a cause of an outcome (Lagnado, Gerstenberg, Zultan 2014; Tadros 2018). Naturally, the debate over **RML** also revolved around this binary question. This is evident in paradigm examples of **RML** in the literature which involve comparing a case in which someone is killed or hurt to one in which no one is killed or hurt. In recent years, however, we've seen an increased interest in the question of *degrees* of causation—that is, roughly, in the relative causal significance of one causal factor over another in bringing about an outcome. A graded notion of causation opens new venues in the debate over **RML** to potentially advance the conversation. Some philosophers have already explored various implications of the idea that causation comes in degrees on **RML**.[41] In this chapter, I'll do the same. But the implications that I'll draw attention to are bad news for **RML**. I'll show that **RML** entails some implausible results that leave **RML** more indefensible than it was previously thought be. These implausible results will also reveal that the purported ordinary moral assessments that support **RML** lead to contradictory results, and hence can't be used as evidence for **RML**. I'll conclude that we should reject **RML**.

---

[40] Cf. Sverdlik (1988:79-80), and Moore (2009:22-23) for discussions on this. Moore says, before he sets out to *defend* **RML**, that "[t]he moral issue is thus better cast in terms other than 'luck.' The issue is better cast straightforwardly in terms of causation: does causation of harm increase moral blameworthiness…?" Cf. Anderson (2019) for an excellent further discussion regarding the role of the notion of "luck" in the moral luck debate.

[41] Cf. Sartorio (2015a), and Bernstein (2017). It should be noted that Sartorio discusses various implications of a graded notion of causation on **RML** without committing herself to the idea that causation comes in degrees. As I noted above, in later work, she argues against graded causation.

Here's one thing to notice before we begin. If, as per **RML**, causal responsibility increases one's degree of moral responsibility, then it's most plausible to hold that more(/less) causal responsibility increases one's degree of moral responsibility more(/less). That is, **RML** is committed to **Proportionality**. One could deny that **RML** is committed to **Proportionality** by holding that causation doesn't comes in degrees.[42] But I already argued that causation comes in degrees. One could also deny that **RML** is committed to **Proportionality** by holding that while causal responsibility is relevant for moral responsibility, *degrees of* causal responsibility isn't. But this would *ad hoc*. It's hard to see why the morally relevant ingredient in causation for responsibility, if there's one, is absent in its degrees.

Notice also that if it's true that **RML** plus graded causation entail **Proportionality**, and since I argued in the previous chapter that **Proportionality** is false, it follows that either **RML** is false or that causation doesn't come in degrees. But I also argued in the first chapter that causation comes in degrees. So now it follows that **RML** is false. Indeed, I'm happy with this result. But now one might worry whether the argument below in this chapter is redundant. Here is why it's not. A proponent of **RML** might want to see my modus tollens as their modus ponens and argue that **Proportionality** is true. So, it's preferable that I further argue against **RML**. And although the implausible results entailed by **RML** that I'll draw attention to below follow also from **Proportionality**, they're not mere quantitative additions to the list implausible results that follow from **RML** or **Proportionality**. Below, I'll also discuss what's new about them.

[42] More precisely, if causation doesn't come in degrees, then **RML** and **Proportionality** are more or less the same thesis.

In the discussion below, various reminders from earlier chapters will again be necessary. As mentioned above—ironically—I'll write as if they occur for the first time since I'd rather not use phrases like "as mentioned in such and such chapter" more times than is necessary.

1. Resultant Moral Luck and Degrees of Causation

Consider these two cases:

> **(Bullet)** Suzy is the only assassin aiming at Victim1. She shoots and Victim1 dies.
> **(Hardy Victim)** Two independently employed assassins, Billy and Lilly, each unaware of the other, are dispatched to eliminate Victim2. Unbeknownst to both assassins, Victim2 is particularly hardy, and requires two bullets for his demise. Each assassin shoots, both bullets arrive at exact same time, and Victim2 dies.

It seems plausible to think that Suzy's causal contribution to Victim1's death is more than Billy's or Lilly's contribution to Victim2's death. This is because while Suzy's input on its own kills Victim1, Billy's or Lilly's input amount to only half of what's needed to kill Victim2. That is, Suzy's degree of causal responsibility for Victim2's death is twice the degree of Billy's or Lilly's degree of causal responsibility for Victim2's.

This is at least the result that we get if we employ a productive theory of causation according to one prominent version of which *c* causes *e* if there's a transfer of conserved quantity (e.g., input, force, or energy) from *c* to *e* (Dowe 2000). For instance, if you hit a glass and it falls off the table, you brought about this outcome in virtue of transferring energy to the glass. Given this account of causation, plausibly, the degree of causation is about how much input a factor transfers to an outcome. Hence, a plausible criterion for measuring degrees of causation is this:

> **(CC1)** The amount of causal contribution to an outcome is the ratio of the amount of transferred input to the amount of input needed for the outcome.

Since it took two bullets for Victim2 to die and Billy and Lilly each shot one bullet, **CC1** tells us that Billy's or Lilly's degree of causal responsibility for Victim2's death is 1/2. And since it took one bullet for Victim1 to die and Suzy's bullet was the only bullet, **CC1** tells us that Suzy's degree of causal responsibility is 1/1. All this is in line with the intuitive idea above that Suzy's causal responsibility is twice the causal responsibility of Billy or Lilly.

Now, recall that resultant moral luck (**RML**) is the idea that all else equal causal responsibility increases one's moral responsibility. And if causal responsibility increases one's moral responsibility, it's most plausible to hold that more(/less) causal responsibility increases morally responsibility more(/less). We then get the conclusion that Billy and Lilly are much less morally responsible than Suzy since their degree of causal responsibility is much less than that of Suzy's.[43] Consider now a further case:

> **(Protection)** A great many assassins, unaware of each other, are set out to kill Victim3. Unbeknownst to them, Victim3 is protected by an electromagnetic field of some sorts. It'll take a great many bullets to kill her. And for all each of them knows, they can kill her with one bullet. They all shoot, and she dies.

In this case, **CC1** tells us that each of these assassins only negligibly contributes to Victim3's death. This is because a great many bullets are needed for her death, and each assassin's input is only one bullet. And **RML** now entails that each of these assassins are only 'negligibly' morally responsible. This should seem worrisome for **RML**. In the next section, I'll discuss in detail what

[43] This is what Bernstein (2017) calls 'Proportionality Luck' as a kind of **RML**. I also adapt (Hardy Victim) from Bernstein (2017). And for those who prefer a characterization of **RML** that emphasizes 'luck,' notice that whether a victim turns out to be hardy or not is out of the assassins' control. Hence, the difference in degrees of causal responsibility between Suzy and, Billy or Lilly, is partly due to 'luck.'

this 'negligible' moral responsibility means. But for now, let's make sure that this result isn't just a quirk of **CC1**.

Consider probabilistic theories of causation according to which causation is a matter of raising the probability of the occurrence of an event (Eells 1991). Hence, a criterion for measuring degrees of causation that follows is this:

> **(CC2)**: The amount of causal contribution is the amount of increased probability, due to a given factor, of the occurrence of an outcome.[44]

Notice that in **(Protection)** each assassin's bullet increases the probability of Victim3's death only negligibly. Hence, we again get the result from **RML** that each of these assassins are only 'negligibly' morally responsible.

I don't want to give the impression that the same result follows no matter what criterion for measuring degrees of causation we apply to **(Protection)**. My contention rather is that even if a criterion doesn't give us the result that each assassin in **(Protection)** only negligibly contributes to the victim's death, we can come up with a different case for that criterion to give us the same result. Consider the following criterion that's based on dependent, or counterfactual, theories of causation:

> **(CC3)** How much a cause contributes to an effect is a matter of how close it comes to providing a necessary condition for an effect.[45]

To illustrate this criterion, consider these two cases:

> **(Bullet)** Suzy is the only assassin aiming at Victim1. She shoots and Victim1 dies.
> **(Victim)** Two independently employed assassins, Timmy and Rosey, unaware of

[44] Cf. Kaiserman (2016) for a more developed version of this criterion.
[45] This is Sartorio's (2020) formulation of the criterion. The same idea can be found also in Bernstein (2017).

> each other, are dispatched to eliminate Victim4. Being struck by one bullet is sufficient to kill Victim4. Each assassin shoots, both bullets arrive at exact same time, and Victim4 dies.

Notice that whether Victim1 lives or dies depends on Suzy's bullet. But Timmy's bullet on its own is far from being a necessary condition for Victim4's death. The same is true of Rosey's bullet. If not for Timmy's bullet, Rosey's bullet would still have killed Victim4, and if not for Rosey's bullet, Timmy's bullet would have killed Victim4. So, **CC3** tells us that the further a factor is from being a necessary condition for an outcome, the less that factor contributes to that outcome. A lot depends here on further details of what it means for a cause to be further from a necessary condition for an outcome. And it may turn out that **CC3** applied to **(Protection)** won't give us the result that each assassin in **(Protection)** only negligibly causally contributes to the victim's death. What's clear, however, is that if there were three assassins in **(Victim)**, **CC3** would tell us that each of those assassins causally contributed to the victim's death even less than how much Timmy or Rosey initially contributed. So, consider now this case:

> **(No Protection)** A great many assassins, unaware of each other, are set out to kill Victim5. It takes only one bullet for him to die. They all shoot, and he dies.

CC3 tells us now that each assassin in **(No Protection)** only negligibly causally contributes to Victim5's death. This is because each of these assassins are extremely far from being a necessary condition for Victim5's death.

I conclude then that the result, in **(Protection)** or **(No Protection)**, that each assassin is only negligibly causally responsible, and hence only negligibly morally

responsible, isn't a quirk of a specific criterion for measuring degrees of causation. Assuming **RML** and graded causation, similar results follow.

2. 'Negligible' Moral Responsibility?

Now, what does it mean for an assassin to be negligibly morally responsible? How morally responsible is it exactly?

One might object that what I called negligible moral responsibility isn't so "negligible." After all, it's morally responsibility for murder. However small degree of moral responsibility for murder it might be, it's still much worse than, say, moral responsibility for breaking a promise, or for lying. So, the assassins may be morally responsible for murder to a least degree. But this is still high degree of moral responsibility.

I think the objection is largely on the right track. But there's still an unanswered question: How high exactly is it to be morally responsible for murder to a least degree? In comparing the moral blameworthiness of someone who merely attempts murder to that of a murderer, Moore—a proponent of **RML**—argues that the former is half as blameworthy as the latter (Moore, 2009:21ff, 2011:505). For instance, if someone who on her own fully contributed to a death is 100 percent blameworthy, someone who merely attempted murder is 50 percent blameworthy. I don't need to commit myself to these exact figures. But the difference must be significant, and not, say, 5 or 10 percent. This suffices for my purposes. If all that follows from **RML** were such small degrees of difference in blameworthiness, **RML** would hardly generate the vast literature that it has so far generated. Moreover, if the difference were so small, intuitive judgements could

hardly play the significant role that it does in the debate over **RML** for it is dubious that our intuitions can track such small differences.

Now if merely attempting and *not* contributing to a death at all makes one 50 percent blameworthy, then plausibly attempting and contributing to a death negligibly makes one negligibly more than 50 percent blameworthy. The objector might then hold that to be morally responsible for murder to a least degree is negligibly higher than 50 percent moral responsibility.

Yet, this isn't exactly good news for **RML**. To see why, recall **(Bullet)** in which Suzy was the only assassin, and compare her to assassins in, say, **(No Protection)**. Given the objector's commitments in the previous paragraph, **RML** now entails that while Suzy is 100 percent morally responsible, assassins in **(No Protection)** are barely over 50 percent morally responsible. But this seems unacceptable. The assassins in **(No Protection)** do kill someone just like Suzy. They went out there and killed someone intentionally and with no excuse. I highly doubt that we'd treat the two sets of assassins differently to such a significant extent. I doubt, for instance, that we'd think that Suzy should receive death penalty or life without parole, while the assassins in **(No Protection)** should receive some years in prison with the possibility of parole. Or I doubt that we'd think that the "report-card of life" (Zimmerman 1985:115) of each assassin in **(No Protection)** receives considerably less demerit than that of Suzy's. It would be unacceptable, to paraphrase Thomson (1993:215), to have God throw assassins in **(No Protection)** into a relatively shallow circle of hell compared to Suzy who would go into a much deeper circle of hell.

3. What's New?

So, **RML** entails some unacceptable results. But the opponents of **RML** think that it's unacceptable to judge, say, reckless or drunk drivers who end up killing someone and reckless or drunk drivers who kill no one. And the proponents of **RML** came to terms with differential moral judgements in these cases. One might then wonder what's new about the unacceptable results above.[46]

In response, first, notice that in paradigm examples of **RML** the comparison is between a case in which someone is killed (or hurt) and one in which no one is killed (or hurt). *If* those who merely attempt murder (or reckless or drunk drivers who don't kill anyone) are morally lucky, this is because they don't end up killing anyone.[47] The proponents of **RML** might then retort that we shouldn't think of those who merely attempted murder as if they killed someone. After all, well, they didn't kill anyone. There is no causal link that ties them to anyone's death. This might seem like a reasonable protest.

As opposed to this, consider comparing, say, **(Protection)** and **(Bullet)**. Notice that **(a)** we're *not* comparing a case in which someone is killed and one in which no one is killed, *and* **(b)** **RML** nonetheless discounts *significantly* from the moral responsibility of murderers in **(Protection)**. That is, **RML** discriminates among murderers who freely and knowingly kill innocent people. [48] The assassins in **(Protection)** do kill someone just like Suzy in **(Bullet)**. Each

---

[46] Thanks to Ben Bradley and Jessica Isserow for pressing me on this.

[47] Compare: "[T]he fortunate reckless driver is morally lucky because he doesn't cause any harm; the unfortunate reckless driver is morally unlucky because he causes harm" (Sartorio 2012:65).

[48] Sverdlik (1988:80-81) discusses the following case. Consider that A shoots, and kills B. Consider also that A wants to kill B, shoots but kills C instead. Notice that, unlike in paradigm example cases of **RML**, this case also involves comparing two scenarios in both of which someone dies. Although Sverdlik thinks that it would commit **RML** to absurd conclusions, he agrees that the more plausible option for **RML** defenders is to say that A is equally blameworthy in both cases, which also straightforwardly follows from my favored characterization of **RML**. So, while Sverdlik's case involves the feature **(a)** of my cases, it doesn't involve the feature **(b)**. That is, Sverdlik's case is *not* a case where **RML** discriminates among the murderers.

of these assassins in both cases is causally linked to someone's death—unlike those who merely attempt murder who are not linked to anyone's death. So, the above reasonable protest now is *not* available to the proponents of **RML**. Hence, discounting from the moral responsibility of assassins in **(Protection)** is more indefensible than discounting from the moral responsibility of those who merely attempt murder, or reckless or drunk drivers who don't kill anyone. It is one thing to hold that unsuccessful attempts to murder is less blameworthy and quite another to hold, in addition, that some murderers are also about as little blameworthy.

One might object that each assassin in **(Protection)** is not really a murderer. This is because, although all of them together kill the victim, each assassin merely non-lethally contributes to the killing. That is, each assassin's contribution on its own wouldn't have killed the victim. In response, first, notice that this objection doesn't apply to **(No Protection)** in which each assassin's contribution *is* lethal. But I'll ignore this. Second, consider this. Billy wants to push Suzy down the cliff yet can't do this on his own. But a large bird flies right into Suzy's head causing her to almost fall down the cliff. Billy seizes this opportunity and, at the same time as the bird hits Suzy, gives her a slight push down the cliff resulting in Suzy's death. Notice that Billy doesn't do the whole work on his own. His slight push is a merely non-lethal contribution—i.e., it wouldn't have killed Suzy on its own. In fact, most of the work is done by the bird. But it's perfectly appropriate to think that Billy is a murderer. Third, consider two people together killing an innocent person by beating the victim. Suppose also that neither of the perpetrators on their own could have killed the victim. Still, it seems perfectly appropriate to think that they're both murderers. Moreover, I doubt that we'd think differently if it were three perpetrators instead. And it seems arbitrary to draw a line somewhere between two people

together killing someone and a great many people together killing someone. Hence, the objection fails. Each assassin in **(Protection)** *is* a murderer.

Second—going back to the "what's new?" question—as mentioned above, the often-used positive argument in favor of **RML** is that our ordinary moral assessments support **RML**. We ordinarily hold, it's argued, that murderers are significantly more morally responsible than those who merely attempt murder. Similarly, we hold that reckless or drunk drivers who kill someone are much more morally responsible than reckless or drunk drivers who don't kill anyone. Now suppose these alleged moral assessments are on the right track. Hence, assume for *reductio* that, say, murderers are significantly more morally responsible than those who merely attempt murder. If this is correct, then we're committed to **RML**. As argued above, **RML** entails the following implausible result: assassins in **(Protection)** or **(No Protection)** are negligibly more than 50 percent morally responsible. And assuming **RML**, this also means assassins in **(Protection)** or **(No Protection)** are negligibly more morally responsible than those who merely attempt murder. Equivalently, it means, some murderers are negligibly more morally responsible than those who attempt murder. That is, it's false that murderers are significantly more morally responsible than those who attempt murder, which contradicts the initial assumption. Hence, we must reject the initial assumption, and along with it the veracity of our alleged ordinary moral assessments. Thus, the implausible results that I reveal in this paper also show that what's typically taken to be the positive argument for **RML** should be rejected.

Here's another implication of the argument in the foregoing paragraph. If sound, it implies that we don't need to wait for empirical studies to settle the question as to whether

people's ordinary moral assessments in question do or don't support **RML**.[49] Even if it turns out that they do, the argument above shows that they can't be used as evidence for **RML**.

4. Conclusion

I revealed some implausible results entailed by resultant moral luck (**RML**). I argued that **(i)** these results make **RML** more indefensible than it was thought to be. I also argued that **(ii)** our alleged ordinary moral assessments that are typically taken to be the evidence for **RML** should be rejected because they lead to contradictory results. I conclude that we should reject **RML**.

Of course, the proponents of **RML** don't employ only the positive argument in question to defend their view. They also employ negative arguments that undermine theories of moral responsibility free of **RML** (Moore 2009:ch.2, Hartman 2016). If these theories are more implausible than accepting **RML**, then maybe all-things-considered we should embrace **RML**. Surely, engaging with these negative arguments in detail would take a longer work. But **(i)** and **(ii)** above leave these arguments dubious. This is because such all-things-considered arguments must now take **(i)** into account, and cannot—as per **(ii)**—derive any weight from the alleged ordinary moral assessments.

[49] As I noted above, at least one study (Kneer and Machery, 2019) shows that they don't.

Chapter 4: Causal Responsibility is Metaphysically Irrelevant for Moral Responsibility

Abstract:

In this chapter, I'll argue that causal responsibility is metaphysically irrelevant for moral responsibility. This is because causal responsibility figures neither in what explains that which one is morally responsible *for* nor in what explains one's *degree* of moral responsibility. And what one is morally responsible for and the degree of moral responsibility are all that needs to be explained about moral responsibility. I will also discuss various theoretical advantages of rejecting causal responsibility as relevant to moral responsibility.

In this chapter I will argue that causal responsibility is metaphysically irrelevant for moral responsibility. This is because causal responsibility does not figure in all that we want to explain about moral responsibility.

When one is apt for praise or blame, it's due to *something* that she is apt for praise or blame. That thing is (part of) what one is morally responsible *for*. Without that thing (or those things) in place, one is not morally responsible to begin with. And of course, we want to know what it is. But *that* one is apt for blame or praise is not all that we want to know. We also want to know *to what extent* one is apt for blame or praise. Put differently, if one is to receive a positive or negative mark in one's "report-card of life" (Zimmerman 1985:115), we want to know its size. That is, we want to know about *the degree* of one's moral responsibility.

To my mind, what one is morally responsible for and the degree of one's moral responsibility are all that we want to explain about moral responsibility. Of course, I don't mean to suggest that there's nothing else to talk about that's relevant to moral responsibility. For

instance, aside from the question of whether one is worthy of praise or blame, there are questions are like whether one actually should all things considered be blamed or praised, or whether there is anyone around for whom it is fitting that they do the blaming or praising. But the relevant considerations for these questions involve much more than the considerations for the worthiness of praise and blame. And my thesis in this chapter (or responsibility internalism itself) is about the *aptness* or one's *worthiness* of praise or blame. Hence, I leave these questions aside.

Below in **§1** through **§2**, I will argue that causal responsibility is metaphysically irrelevant respectively to the *degree* of moral responsibility and to what one is morally responsible *for*. I will conclude that causal responsibility is metaphysically irrelevant for moral responsibility. In **§3**, I will discuss various theoretical advantages of this conclusion.

1. The Degree of Moral Responsibility

In previous two chapters, I argued that causal responsibility neither increases nor is proportionate to moral responsibility. To my knowledge, no one in the literature holds any other view regarding the relationship between causal responsibility and the degree of moral responsibility. Nonetheless, the logical space allows roughly two more options. In this section, I will argue that none of them are viable options and conclude that causal responsibility is metaphysically irrelevant for the degree of moral responsibility.

First, rather than thinking that causal responsibility in some specific way *increases* the degree of moral responsibility, one might hold that it somehow *decreases* the degree of moral

responsibility. But unless such an option has some initial plausibility, it's safe to ignore it. And it's hard to see how such an option can even be motivated.

Second, rather than thinking that causal responsibility increases the degree of moral responsibility full stop, one might hold that causal responsibility *sometimes, but not always,* increases the degree of moral responsibility. This thesis could be understood as either involving a *ceteris paribus* clause or not involving it.

The former thesis tells us that holding fixed the exercise of control, epistemic status, intentions, motivations, cares, and choices, causal responsibility only sometimes increases the degree of moral responsibility. But it's unclear why causal responsibility *only* sometimes has such an effect on moral responsibility. That is, despite the fact that nothing else that's relevant for moral responsibility changes, causal responsibility sometimes becomes morally relevant and sometimes not. This is no different than saying that the moral value of causal responsibility randomly goes in and out of existence. But this seems implausible.

The latter thesis tells us that causal responsibility can increase the degree of moral responsibility when all else is *not* equal. For instance, one might think that when one exercises a stronger degree of control over, or 'knows' or cares more about, one's choice of action, one's causal responsibility increases one's degree of moral responsibility. But this is dubious. If there's an increase in one's degree of moral responsibility when one exercises a stronger degree of control over, or 'knows' or cares more about, one's choice of action, it's unclear why we need to appeal to one's causal responsibility *in addition to* one's degree of control, awareness, or care in order to explain one's degrees of moral responsibility. It gets even more

unclear why we need to do so once we consider, as shown in the previous paragraph, that causal responsibility on its own doesn't affect the degree of one's moral responsibility. The thesis would be telling us that while causal responsibility lacks inherent relevance to degrees of moral responsibility, it gains such relevance when it's combined with factors that are inherently relevant to one's degree of moral responsibility. I suspect there's nothing obviously inconsistent with this suggestion, but it's unclear why it's needed or how it could be motivated.

To illustrate, consider the following cases:

> **(Punch-to-Kill)** Suzy wants to kill Sally. She knows that if she punches Sally in the face, Sally will die. Suzy punches Sally in the face and Sally dies.
>
> **(Punch-to-Wound)** Timmy wants to wound Terry by punching him in the face—a punch that's on a par to Suzy's punch to Sally. For all he knows, the punch will only wound Terry. He punches Terry, and Terry gets wounded.
>
> **(Punch-to-Wound*)** Billy wants to wound Barry by punching him in the face—a punch that's on a par to Suzy's punch to Sally. For all he knows, the punch will only wound Barry—just like for all Timmy knew, his punch would only wound Terry. Unbeknownst Billy, however, the punch will kill Barry. Billy punches Barry, and Barry dies. In all other relevant respects, Billy and Barry are on a par.

It seems that Billy is less blameworthy than Suzy. In fact, I think, while Suzy is as blameworthy as a murderer, Billy is as blameworthy as Timmy who is much less blameworthy than Suzy. One might think that Billy must be *more* blameworthy than Timmy—though he plausibly is less blameworthy than Suzy. But this amounts to holding that *all else equal* causal responsibility increases one's degree of moral responsibility, and I argued that this is false. And to explain why Suzy is as blameworthy as a murderer and Billy is as blameworthy as Timmy, we need only to appeal to the differences between what Suzy or Billy is aware of, their motivations or what they

intended to do. It's unclear what appealing to their causal responsibilities could add to the explanation.[50]

To sum it up, causal responsibility neither increases, nor is proportionate to, nor decreases, the degree of moral responsibility. It's also false that causal responsibility sometimes, but not always, increases the degree of moral responsibility. Hence, I conclude that causal responsibility is metaphysically irrelevant to the degree of moral responsibility.

2. Moral Responsibility *For*

We might mean two things by "S is morally responsible *for* Φ." One, "S is morally responsible (partly) *because of*, or *in virtue of*, Φ" or "Φ is (partly) what *makes* S morally responsible." Two, "S is morally responsible *to* Φ," "S *has* a responsibility regarding Φ," or "S is to *take* responsibility for Φ." In this section, I'll argue that causal responsibility is metaphysically irrelevant for the former, and although it might be relevant for the latter, this is benign for my purposes.

Here are my two reasons why causal responsibility is metaphysically irrelevant for the former. One, a satisfactory account of that which makes one morally responsible can be given without an appeal to causal responsibility. Hence, causal responsibility is redundant in explaining what makes one morally responsible. Two, it's implausible to think that while causal responsibility is irrelevant for the degree of moral responsibility, it is (part of) that which makes

[50] Notice that in this paragraph I merely illustrate my argument in the previous paragraph. So, my contention isn't that the degree of moral responsibility can be explained solely by what one is aware of or what one's intentions or motivations are, or that the explanation always involves these factors. For instance, some might think that not only what one is aware of, or not only one's intentions or motivations, but *the lack thereof* might be relevant to the explanation (e.g., in cases of negligence). This is consistent with my view. My contention ultimately is that causal responsibility isn't needed to explain the degree of moral responsibility.

one morally responsible. This is because it commits one to hold that the degree to which one is responsible can be determined before (part of) what makes one morally responsible is fixed.

To illustrate, consider the following cases:

> **(Throw-and-Hit)** Suzy wants to wound Sally by throwing a rock at her. She throws a rock, the rock hits Sally, and Sally gets wounded.
>
> **(Throw-and-Miss)** Billy wants to wound Barry by throwing a rock at him. He throws a rock at him but misses the target due to a strong gust of wind changing the trajectory of the rock.
>
> **(Throw-Attempt)** Timmy wants to wound Terry by throwing a rock at him. He decides to throw a rock at him, sets out sets out to move his arm to throw the rock, but a strong gust of wind pushes his arm back rendering him unable to throw the rock.

Suzy, Billy, and Terry are all morally blameworthy. They are also blameworthy to the same degree because they differ only in their causal responsibilities and causal responsibility is irrelevant to the degree of moral responsibility. But what makes them morally blameworthy? It is clear that Terry is not morally blameworthy in virtue of his causal responsibility because he is not causally responsible for anything. He wills, tries, or sets out to cause something, but he fails. So, a satisfactory account of what makes him blameworthy might include factors such as his exercise of control over his decision and setting out to hurt Terry, his awareness of various relevant factors or lack thereof, and his motivations, intentions, cares or lack thereof, but it excludes his causal responsibility. If this is correct for Timmy, it's unclear why it should be any different for Billy or Suzy.

Suppose we want to say that there is more involved in making Billy or Suzy morally blameworthy. So, what makes Billy morally blameworthy, in addition to what makes Timmy blameworthy, is that he is causally responsible for throwing a rock. What makes Suzy morally

blameworthy, in addition to what makes Timmy blameworthy, is that she is causally responsible for throwing a rock *and* for wounding Sally. Notice that we could come up with more elaborate stories in which one is causally responsible for, and hence purportedly morally responsible in virtue of, even further things—numerous further things. But it's unclear why we should thus keep overpopulating the ground on which one is morally blameworthy. It is especially unclear once we consider that all these things that we keep adding to the ground that makes one blameworthy leave the degree to which one is blameworthy untouched. Hence, it seems, causal responsibility is redundant in explaining what makes one morally responsible.

Moreover, notice that once they set out to throw a rock, the degree of their blameworthiness is fixed and Suzy, Billy, and Timmy are blameworthy to the same extent. This is because the rest of what happens belongs to their causal responsibilities (or lack thereof) and I argued that causal responsibility is irrelevant for the degree of moral responsibility. So now if one wants to hold that, say, what makes Suzy morally responsible is also her causal responsibility for hurting Sally, one is committed to thinking that Suzy's degree of moral responsibility is fixed even before what makes Sally morally responsible is fully in place. But this seems odd. It seems implausible to think that *that* one is morally responsible comes temporally after one's *degree* of moral responsibility is fixed. By analogy consider the idea that the strength of a friendship can be fixed and in place even before what grounds that friendship, and hence that friendship itself, is in place, which seems quite odd. Hence, again, it seems implausible to hold that causal responsibility is (part of) what makes one morally responsible.

One might worry that in some cases where one seems to be blameworthy there isn't any factor other than one's causal responsibility that's readily available to make one

blameworthy. Consider a drunk driver who gets into an accident. At the time of the accident, she's not in control of what she's doing and it's hard to attribute her any relevant awareness, intention, or care. But nonetheless she seems blameworthy. In response, it is true that *at the time of the accident* the drunk driver lacks control, awareness, intention, or care. Notice, however, that it might be true also of Suzy *at the time of rock's hitting Sally* that Suzy lacks all these features. Consider this:

> **(Throw-and-Hit*)** Right after Suzy throws the rock and before the rock hits Sally, Suzy is rendered unconscious by someone punching her in the head. Everything else is the same as **(Throw-and-Hit)**.

Surely, Suzy is blameworthy. And we have no reason to think that what makes her blameworthy is any different than that which made her blameworthy in **(Throw-and-Hit)**. So, in **(Throw-and-Hit)** whether Suzy lacks control, awareness, intentions, and care at the time of the rock's hitting Sally is irrelevant to whether Suzy is blameworthy or what makes her blameworthy. Similarly, whether the drunk driver lacks all these features at the time of the accident is irrelevant to whether she's blameworthy or what makes her blameworthy. All we need to do to determine whether she's blameworthy or what makes her blameworthy is to trace things back to where she has all these features just like we do for Suzy in **(Throw-and-Hit*)**.

One might worry that there is a remaining moral difference between **(Throw-and-Hit)** on one hand, and **(Throw-and-Miss)** and **(Throw-Attempt)** on the other. After all, Suzy did wound someone whereas Billy or Timmy didn't. And it seems appropriate that Suzy now has a duty of reparation or should somehow make amends with Sally whereas no further action is required of Billy or Timmy because they didn't wound anyone.

This brings us to the second sense of "S is morally responsible *for* Φ" which again is that "S is morally responsible *to* Φ," "S *has* a responsibility regarding Φ," or "S is to *take* responsibility for Φ." The above concern is that while Suzy should *take* responsibility for wounding someone, Billy or Timmy need not take responsibility for anything. So, in this sense of being morally responsible *for*, Suzy is morally responsible for more things Billy or Timmy. In response, notice that now we are *not* talking about moral responsibility in the basic desert sense—in the sense of being apt for praise or blame. We're talking about moral responsibility in the duty sense: what is it that Suzy, Billy, or Timmy should do? And my main thesis in this chapter—that causal responsibility is metaphysically irrelevant for moral responsibility—is about the former. So, even if the above concern is on the right track, it's consistent with my view that Suzy is morally responsible for more things, i.e., she has more moral duties, compared to Billy and Timmy.

Indeed, *that* two views are consistent with another doesn't mean that the combination of those views add up to a plausible view. So, one might worry whether I'm committed to an implausible view now. And here are two major reasons why one might have such a worry. One, I argued that Suzy's causal responsibility, which also involves wounding Sally, isn't part of what *makes* Suzy blameworthy. So, one might now worry whether it's plausible to hold that Suzy has a duty to make amends with Sally for wounding her despite the fact that wounding her isn't part of what makes Suzy blameworthy. Two, if Suzy has an extra duty while Billy or Timmy doesn't, despite the fact that they are all blameworthy on the same grounds, one might worry that this is unfair to Suzy.

Let's begin by noting that being blameworthy for something is not necessary for a moral duty to incur regarding that thing. First, consider moral duties like telling the truth, keeping one's promises, being non-maleficent, or helping those in need. None of them requires being blameworthy for anything. Second, consider that *mere* causal responsibility is often sufficient to generate a duty. Suppose while walking home, you get caught up in a heavy rainstorm which causes very low visibility. You want to avoid getting wet and sick. So, you start walking fast. Shortly after, you run into someone—who's also walking fast to avoid getting wet—and knock him down. You're causally responsible for what happened to him, but you're not blameworthy for it given the circumstances. What happened is an accident. But surely you have a duty now—you should help him up and get cleaned. Or suppose you take someone else's property not knowing that it belongs to someone else. When it turns out later that it does belong to someone else, you have a duty to give it back, which doesn't require you to be blameworthy for taking it in the first place. Third, consider cases where, *through no fault of your own*, your teenage kid, pet, or farm animal causes harm to others. You're not blameworthy for the harm, but it's perfectly plausible that it's your duty to compensate for it. So, being blameworthy for something is *not* necessary for having a duty concerning that thing. Indeed, holding otherwise commits one to implausible claims.

Sure, it doesn't follow that Suzy doesn't have a duty to make amends with Sally. But it does follow that *if* Suzy has a duty, it's dubious that this is because she's blameworthy for hurting Sally. It's entirely open to hold that Suzy's duty arises from her being morally blameworthy for deciding, and setting out, or willing, to harm Sally *plus* her being causally responsible for hurting Sally (Khoury 2018). Or it could be that Suzy's duty arises from what

Susan Wolf (2001) calls the *nameless* virtue—the virtue of *taking* responsibility for the consequences of one's actions even if those consequences are unforeseen, unintended, or somehow out of one's control. One might still worry that the foregoing two suggestions are a little 'forced.' But, contrary to appearances and as the above discussion shows, there isn't a simple, straightforward connection between the two senses of responsibility—i.e., being blameworthy or praiseworthy versus having duties.[51]

Let's now turn to the second worry above which was this: if Suzy has an extra duty while Billy or Timmy doesn't, despite the fact that they are all blameworthy on the same grounds, one might worry that this is unfair to Suzy. First thing to note is that I'm not committed to holding that no new duties arise for Billy or Timmy. I suspect most blameworthiness might entail at least one specific duty—roughly, a duty to revise oneself, to try and be better. It's plausible to think then that a new moral duty incurs also for Billy or Timmy. Of course, this duty would apply also to Suzy *besides* her duty to make amends with Sally. So, one might push the point that maybe it's unfair that Suzy has this additional duty while Billy or Timmy doesn't. After all, it is merely a random gust of wind that generates this difference. However, notice that this isn't problematic for duty sense of moral responsibility. A random gust of wind *can* make a difference to incurring of a moral duty. If a gust of wind pushes a little kid into a pond, and

[51] In this regard, it's worth noting that blameworthiness for something may also not be sufficient to generate a duty concerning that thing. If you're blameworthy for hurting someone, but another person nearby is much more in need of a help, you may have a duty to help that other person instead. And even if being blameworthy for something does generate a duty regarding that thing, the exact content of that duty is still a further question. Suppose you're the blameworthy driver in a car crash where your car and another car are wrecked. It doesn't yet follow that your duty is to compensate for the other car's cost. Suppose the owner of the other car is extremely rich while you barely make the ends meet. It seems (morally) pointless to assign you the duty for compensation. Surely you should at least apologize to the other driver. But that's precisely the point—the exact content of your duty, or what your duty is, isn't determined only by that which you're blameworthy for.

you're the only one there to help, a new duty arises for you to save her life. Indeed, incurring of most of our moral duties are at least partly due to factors we cannot control if not in fact all of them. If someone happens to ask you a question, you have a (prima facie) duty to tell the truth. If someone happens to be around you, you have a (prima facie) duty to be non-maleficent towards them. The point is that unless one is prepared to call all these duties unfair or implausible, one shouldn't think of Suzy's duty to make amends with Sally unfair or implausible.

I conclude that causal responsibility is metaphysically irrelevant to what one is morally responsible *for*—responsibility understood in the basic desert sense. It might be relevant to what one is responsible *for*—responsibility understood in the duty sense. But this is neither inconsistent with my view nor makes my view an implausible one. I also argued in the previous section that causal responsibility is metaphysically irrelevant to the degree of basic desert responsibility. Hence, I conclude, causal responsibility is metaphysically irrelevant for basic desert responsibility.

3. Various Advantages of My View

Needless to say, I think my argument above is sound. But the thesis that causal responsibility is metaphysically irrelevant for moral responsibility is also theoretically very advantageous. In this section, I will discuss three main advantages of this thesis.

One, consider Principle of Alternate Possibilities (**PAP**) according to which availability of alternate courses of action is a necessary condition for moral responsibility. To illustrate, suppose an assassin kills a victim. The principle tells us that the assassin is blameworthy only if she could have done otherwise—that is, roughly, only if she could have refrained from killing

the victim. **PAP** played, and continues to play, a central role in the debate over the control condition for moral responsibility—i.e., for free will. Compatibilists about free will typically hold that **PAP** is false. In fact, plausibly, compatibilists are committed to thinking that **PAP** is false. But many libertarians—i.e., incompatibilist defenders of free will—also think that **PAP** is false.[52] So, plausibly, it's widely accepted that **PAP** is false.

Notice that if causal responsibility is irrelevant for moral responsibility, then whether one ends up bringing about one thing or another is irrelevant for moral responsibility. That is, whether one follows one course of action or another among the ones that are available to him in future is irrelevant for moral responsibility. It follows that **PAP** is false because the availability of alternate courses of action in future is irrelevant for moral responsibility. Hence, the thesis that causal responsibility is irrelevant for moral responsibility gives us a good explanation for a widely held view—that **PAP** is false.[53]

Two, if causal responsibility *is* relevant for moral responsibility, then the correct moral diagnosis of a given case is hostage to the correct causal diagnosis of that case. And the question regarding the correct causal diagnosis can get very messy. To illustrate, consider the perennial puzzle of thirsty traveler.[54] Here's a basic version of the thought experiment that gives rise to the puzzle:

> **(Thirsty Traveler)** Billy and Suzy, unbeknownst to one another, want Sally dead. Sally, the traveler, has a canteen full of water that she will need to drink to survive. To kill

[52] See, e.g., Stump (1999), Hunt (2000), Zagzebski (2000), and Shabo (2010). Some non-libertarian incompatibilists—i.e., responsibility skeptics—also reject **PAP** (Pereboom 2014).
[53] Zimmerman (2006:601-3) makes a similar point.
[54] The original thirsty traveler case is from MacLaughlin (1925). Cf. Kvart (2002), Sartorio (2007, 2015b), Moore (2009:466-7), Bernstein (2019) for further discussion on various causal and moral diagnoses for this case. Cf. Alexander (2021, pp.365-7) for further examples of the kinds of cases where causal diagnosis might get messy.

> Sally, Billy fills the canteen with a poison that kills by dehydration. Afterwards, unaware of what Billy did, Suzy steals Sally's canteen to kill her. When Sally gets thirsty, she reaches out for her canteen but can't find it. Soon after, she dies of dehydration.

It's difficult to give a satisfactory causal diagnosis for this case. It's hard to think that Billy killed Sally because it's not Billy's poison that killed her. It's also hard to think that Suzy killed her because preventing someone from drinking poisoned water doesn't kill them. And it's equally hard to think that they didn't kill her because if not for what they did she wouldn't have died. But if neither Suzy nor Billy is causally responsible for her death, it's difficult to see how Suzy and Billy conjunctively or disjunctively could be causally responsible for her death. Out of nothing, comes nothing.

If one holds that causal responsibility *is* relevant for moral responsibility, one needs to solve this puzzle before one can offer a moral diagnosis for the case. And that's one good thing about the thesis that causal responsibility is *not* relevant for moral responsibility. We don't need to wait for the correct theory of causation to arrive before we can morally assess this case. Moreover, I think, this thesis gives us the intuitively correct result which is not guaranteed on the assumption that causal responsibility *is* relevant: Billy and Suzy both *are* morally blameworthy, and they each are as blameworthy as a murderer. It's difficult to think that they're *not* blameworthy or somehow (much) *less* blameworthy than a murderer.

Three, some philosophers think that we need an account of luck to determine (whether there is moral luck, and by extension) whether there is resultant moral luck. This is because they prefer to define resultant moral luck as follows: resultant moral luck occurs only when luck

in the results of an action increases one's degree of morally responsibility. Hence, if we don't know what result is or isn't lucky, we can't determine whether there is resultant moral luck.

What counts as lucky ranges from 'all results' all the way to 'no results,' and in the literature we find proponents of both ends of the spectrum as well as proponents of various middle ground positions.[55] It is unlikely that the debate will settle anytime soon if at all. It is telling to recall the fate of the search for a non-accidentality—i.e., anti-luck—condition for accounts of knowledge in the post-Gettier literature. Hence, we shouldn't want the correct assessment of moral responsibility to be hostage to the correct account of luck—*if* there is such an account.[56] And the thesis that causal responsibility is irrelevant for moral responsibility grants us just what we should want. If true, the thesis entails that whether a result of an action is lucky or unlucky is irrelevant for moral responsibility.

4. Conclusion

I argued that causal responsibility is metaphysically irrelevant to **(a)** what one is morally responsible *for* and to **(b)** the *degree* of moral responsibility. And since **(a)** and **(b)** are all that needs to be explained about moral responsibility, I concluded that causal responsibility is metaphysically irrelevant for moral responsibility. I also showed that this conclusion comes with significant theoretical advantages. It provides an explanation for a widely held view—that the

[55] Alexander (2021:364) says "[o]nce one acts, the results… are a matter of luck." Hales (2015) is skeptical about all accounts of luck, and hence suggests that there's no problem of moral luck to begin with. Cf. Levy (2011), Whittington (2014), and Peels (2015) for various other accounts of luck and ensuing discussions over moral luck.
[56] As I mentioned in the previous chapter, it's also dubious that emphasizing the role of luck in defining (resultant) moral luck or in the debate over (resultant) moral luck is any fruitful. Cf. Hartman (2017:23-31) and Anderson (2019) for further discussion on this.

Principle of Alternate Possibilities is false—and allows us to ignore various philosophically messy debates in assessing moral responsibility.

Chapter 5: The Epistemic Condition and The Control Condition for Moral Responsibility

Abstract:
In this chapter, I will discuss two of the conditions for moral responsibility: the epistemic condition and the control condition. I will argue that moral responsibility does not require any form of true belief. Hence, the epistemic condition does not require anything external to agents. I will also argue that the leading theories of responsibility relevant control, in their most plausible forms, don't require anything external to agents. Hence, the control condition doesn't require anything external to agents. I will conclude that the control condition and the epistemic condition for moral responsibility depend only on factors internal to agents.

Here is where we are at in my argument for responsibility internalism. I said that there are three potential conditions for moral responsibility: the control (or freedom) condition, the epistemic (or awareness) condition, and the causal responsibility condition (or consequences). And so far, I argued that causal responsibility is metaphysically irrelevant for moral responsibility. In this chapter, I will argue that the other two conditions—the epistemic condition and the control condition—depend only on factors internal to agents.

Let me begin again by briefly illustrating these two conditions for moral responsibility. Suppose you push a button thinking that it will only turn on the light in your room. Unbeknownst to you the button is connected to a mechanism in the next room to torture an innocent person. So, you end up torturing an innocent person which plausibly is morally wrong. But intuitively it doesn't seem like you're blameworthy—assuming, of course, that you're not culpable for your unawareness of the person in the next room. Moreover, if you didn't push the

button freely, you're again not morally responsible for what you did. For instance, if you were under hypnosis when you pushed the button, intuitively you're not blameworthy.

1. The Epistemic Condition Doesn't Depend on Factors External to Agents

As illustrated above, moral responsibility requires an awareness of certain relevant factors. Two questions arise now. One, what are these relevant factors? Two, what kind of awareness is required? Potentially relevant factors involve awareness of one's action, the consequences of one's action, alternative courses of action, and perhaps the moral significance of these factors. And the awareness in question could be knowledge, justified belief, true belief, or belief.

So, the potentially relevant factors involve elements external to agents—such as the consequences of an action, or the moral significance of those consequences. Coupled with certain kinds of awareness, this will threaten internalism. For instance, if responsibility requires *knowledge* of the consequences of one's action, then internalism is false. This is because knowledge requires truth and truth of elements external to agents requires something external to agents. Consider, for example, the correspondence between the belief and the extramental element.

However, as many philosophers believe, knowledge is too strict a requirement and hence isn't necessary for moral responsibility (Baron 2017: 58-9, Haji 2008:90, Peels 2014:493-4, Rosen 2008:596, Zimmerman 1997:412). Consider the following case discussed by Rosen (2008:596):

> Dorfman poisons Mrs. Dorfman by putting what he takes to be arsenic in her tea. The stuff is indeed arsenic and Mrs. Dorfman dies as planned. But Dorfman does not know that the stuff is arsenic (or that his act subjects his victim to an unjustifiable risk of death) because: The chemist who sold Dorfman the arsenic is a famous liar … [G]iven the chemist's well-known track record of selling sugar as arsenic to would-be poisoners, Dorfman had no business believing him. Dorfman's pertinent beliefs are true, but they do not amount to knowledge because they are based on insufficient evidence.

Intuitively Dorfman is blameworthy for what he did. It's implausible to think that he's not blameworthy because he didn't *know* that he was poisoning his wife or that his wife would consequently die. Hence, the kind of awareness required for moral responsibility isn't knowledge. Notice also that Dorfman doesn't have a *justified* belief. His belief that he has arsenic is based on insufficient evidence, and on an unreliable belief forming process—the chemist is a liar. This suggests that the kind of awareness required for responsibility isn't justified belief either.[57]

But Dorfman does have a *true* belief. This might suggest that true belief is necessary for moral responsibility. And again, coupled with certain potentially relevant factors for moral responsibility, this will threaten internalism. For instance, if Dorfman must have a true belief about the consequences of his action for him to be morally responsible, then internalism is false. This is because the truth of this belief requires an element external to Dorfman—i.e., his wife's death.

[57] It's clear that if reliabilism about epistemic justification is true, Dorfman's belief is not justified. (See Goldman and Beddor (2021) for a discussion on reliabilism.) Similar cases could be created to satisfy other externalist theories of epistemic justification. But if internalism about epistemic justification is true, Dorfman's belief may or may not be justified. However, this is benign for my purposes because internalism about epistemic justification isn't a threat to responsibility internalism. (See Pappas (2017) for a discussion on externalist and internalist theories of justification.)

However, *true* belief about the potentially relevant factors external to agents is *not* necessary for moral responsibility. Here's an argument for this. I will call it the Argument from Moral Encouragement. Consider a significant aspect of our moral practices—namely, moral encouragement. We typically encourage people—ourselves included—to live a moral life which involves encouraging people to seek the morally right courses of action and act accordingly with one's findings. Of course, it's undesirable that we go around pestering people to live a moral life. But encouraging people for a moral life is typically our default position, and we do encourage people for a moral life if the right opportunity arises. It's largely because we so encourage people that we also promote thinking and theorizing about the correct moral principles. We want to find out what's morally right or wrong and act accordingly with our best findings. It would certainly be strange, moreover, to think that we have no moral reasons for moral encouragement. So, not only moral encouragement is a significant aspect of our moral practices, but we also have good moral reasons for it.

Yet, despite our earnest search for the right courses of action, our findings could be mistaken in all potentially morally relevant respects. We could be mistaken about the moral significance of our actions, their consequences, or alternative courses of action. Suppose, for instance, Suzy is facing a moral dilemma. She earnestly searches what she should do and even consults our best moral theories. Finally, she forms her beliefs and performs the action that she thought was right. As it turns out, however, her beliefs are false, and she performed a morally wrong action. We could hardly find her blameworthy. In fact, we should find her praiseworthy because we do, as we should, encourage her to do exactly as she did. And if we encourage someone to do something, and they do exactly as we encourage them to do, we should only

praise them for it. This suggests that one can be praiseworthy for an action even if one's beliefs regarding the potentially morally relevant factors external to one are all false.

We can also run a parallel argument for blameworthiness. Suppose that Suzy performs a morally right action which, after all her pondering and research as above, she believed was wrong. We should find her blameworthy because she did the exact opposite of what we encourage her to do. She did what we *discourage*, or seek to prevent, her from doing. And if we discourage someone from performing an action and they perform it anyway, we should only blame them for it. This suggests that one can be blameworthy even if one's beliefs regarding the potentially morally relevant factors external to one are all false.

One might object that just because we should praise Suzy doesn't mean she's in fact praiseworthy or that just because we should blame her doesn't mean that she's blameworthy. This is because there might be other reasons (e.g., instrumentalist reasons, or reasons due to a threat) to praise or blame someone even when they don't deserve praise or blame. In response, it would be odd to think that moral encouragement—a significant aspect of our moral practices for which we seem to have good moral reasons—leads us to praise people while they're in fact not praiseworthy and to blame them while they're in fact not blameworthy. We would have to hold that moral encouragement leads us to pretentious moral practices—that we should behave *as if* they are praiseworthy while believing that they are not and that we should behave *as if* they are blameworthy while believing that they are not. We would also have to hold that we sometimes should encourage people to act in *un*praiseworthy ways, and discourage people from acting in *un*blameworthy ways.

Alternatively, an objector might think that our practice of moral encouragement is flawed and hence should be given up. But it would be too costly to give up encouraging people to strive for a moral life. Not only would we have to stop encouraging people—including ourselves—to seek what's morally right or wrong and act accordingly with the best findings, but we'd also lose a significant motivation in our collective endeavor for theorizing about the correct moral principles. Imagine telling people that our best moral theories tell them to perform a certain act, but nonetheless they may not be praiseworthy for it.

An objector might point out that we encourage people to do the right thing, and not merely to seek what's right and act accordingly with the best moral findings.[58] In response, encouraging people to do the right thing, in practice, hardly amounts to more than encouraging them to seek what's right and act accordingly with the best moral findings. To illustrate, imagine that someone who's lost about what they should do asks you for advice and your response to them is "You should do the right thing." This is hardly helpful because it adds virtually nothing to the conversation.[59] The sensible thing to do to further the conversation is to tell them what that right thing is as you believe if you're already prepared to do so. If not, you should encourage them to seek what's morally right, maybe also suggest that you do this together. Either way, you end up encouraging them to act accordingly with the best moral findings.

[58] Thanks to Stephen Kershnar for raising this concern.
[59] Notice that "You should do the right thing" is akin to a tautology. This is because "You should do X" entails that "X is the right thing to do."

So, the Argument from Moral Encouragement suggests that *beliefs that one has* need not be true for one to be morally responsible. For the purposes of moral responsibility and regarding the beliefs that one has, what matters is what one takes to be the case and not what is in fact the case. However, some philosophers argue that not only beliefs that one has but also some *beliefs that one should have* are relevant for moral responsibility. For instance, sometimes when people perform unwitting acts, we still find them blameworthy because 'they should have known better.' Consider the following case discussed by Sher (2009:24):

> **(Hot Dog)** Alessandra, a soccer mom, has gone to pick up her children at their elementary school. As usual, Alessandra is accompanied by the family's border collie, Bathsheba, who rides in the back of the van. Although it is very hot, the pick-up has never taken long, so Alessandra leaves Sheba in the van while she goes to gather her children. This time, however, Alessandra is greeted by a tangled tale of misbehavior, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out. During that time, Sheba languishes, forgotten, in the locked car. When Alessandra and her children finally make it to the parking lot, they find Sheba unconscious from heat prostration.

Notice that Alessandra is unaware that she left Sheba in the car, and hence unaware that she's failing to do something she should do. Despite this fact, some philosophers hold that Alessandra is blameworthy because she could and should have been aware of the relevant factors.[60] This seems to suggest that she is blameworthy because of certain possible beliefs that are true—e.g., that she in fact left Sheba in the car. If this is right, then a true belief can be relevant to moral responsibility. Hence, the awareness requirement for moral responsibility doesn't depend solely on factors internal to agents.

[60] See, e.g., Clarke (2017), Fitzpatrick (2017), Rudy-Hiler (2017), and Sher (2009). However, some philosophers hold that one cannot be *directly* morally responsible for such unwitting or negligent acts or omissions, or that one can't be responsible for such acts or omissions at all. See, e.g., Kershnar (2018:ch.7), Levy (2017), Rosen (2004), and Zimmerman (1997).

In response, first, even if one could be morally responsible because of what one could and should have been aware of, it doesn't yet follow that the awareness in question is of what is in fact the case. It is consistent to hold that one can be morally responsible because of what should have been aware of, but that the awareness in question is of what one should have taken to be the case. And what one should have taken to be the case doesn't presuppose what is in fact the case—i.e., truth. Second, if one insists that the awareness in question is of what is in fact the case, one has to explain the following asymmetry. As I argued above, the relevant beliefs that one *has* don't have to be true for one to be morally responsible. But the objector now asserts that the beliefs that one *should have* requires an extra element—i.e., truth. Third, and to also bolster the previous two points, consider a variant of the Hot Dog case:

> **(Hot-Dog-False-Belief)** Everything is the same as the Hot Dog case except that Alessandra had a *false* belief that she left Sheba in the van. As she's leaving the van, for all she knows she just left Sheba in the van. But Sheba is actually *not* in the van.

Intuitively, *if* Alessandra is blameworthy in **(Hot Dog)**, she's blameworthy in **(Hot-Dog-False-Belief)** just the same. One might object that she is *not* blameworthy in the latter case because, unlike in the former, Sheba didn't in fact suffer. But if, as I argued, causal responsibility or consequences are irrelevant for moral responsibility, this objection fails.[61] And if Alessandra is blameworthy in the latter case, this could only be due to the *false* belief that she left Sheba in the car. This is because this false belief that she had but forgotten is the only other difference

[61] Notice also that thinking that Alessandra is *not* blameworthy in **(Hot-Dog-False-Belief)** would commit one to holding that one cannot be blameworthy unless some unwanted results follow from one's actions. But even those who think that causal responsibility *is* relevant to moral responsibility, or the proponents of resultant moral luck, wouldn't accept this. Recall, for instance, that in comparing murder and attempted murder, they wouldn't claim that merely attempting to murder doesn't make one blameworthy at all. All they hold is that murderers are *more* blameworthy than those who merely attempt murder. So, one doesn't have to hold that causal responsibility is irrelevant for moral responsibility to hold that Alessandra *is* blameworthy in **(Hot-Dog-False-Belief)**.

between the two cases. This suggests that if Alessandra is morally responsible, it's because of what she should have taken to be the case, and not because she should have been aware of what is in fact the case. And if this is the correct explanation of why Alessandra is blameworthy in the latter case, it should also be the correct explanation of why she is blameworthy in the former case. Hence, even if one can be morally responsible because of what one should have been aware of, the awareness in question is *not* of what is in fact the case but of what one should have taken to be the case.

To sum it up, the awareness requirement for moral responsibility could be knowledge, justified belief, true belief, or belief concerning potentially responsibility relevant factors. I argued that this awareness can't be knowledge, justified belief, or true belief. The awareness in question is not of what's actually the case but of what the agent takes (or should have taken) to be the case. And what one takes to be the case doesn't presuppose any elements external to agents. Hence, the awareness requirement for moral responsibility depends only on factors internal to agents.

2. The Control Condition Doesn't Depend on Factors External to Agents

In this section, I'll present the leading theories of responsibility relevant control and argue that none of them, in their most plausible forms at least, requires external elements.

As mentioned above, the control in question is the subject matter of the classic free will debate. Broadly speaking, there are two kinds of theories of control: compatibilist theories and incompatibilist theories. The two leading compatibilist theories are the deep-self view and the

guidance control theory.[62] According to the deep-self view, one has responsibility relevant control just in case one's behavior is controlled by one's deep self. One's deep self, in turn, is defined in terms of one's second-order volitions, cares, commitments, values, perceptions, or judgements of the good.[63] For instance, on Frankfurt's (1971) account, one has responsibility relevant control just in case one's act is governed by one's will and one's will is governed by his second-order volitions. A second-order volition is a second-order desire that a first-order desire be a *will* (that is, a desire that brings about an act). On Watson's (1975) view, one has responsibility relevant control just in case one's action is governed by a value, where a value functions similar to a second-order volition.

According to the guidance control theory, one has responsibility relevant control just in case one's action issues from one's own moderately reasons-responsive mechanism (Fischer 1994, 2007; Fischer and Ravizza 1998). A mechanism is moderately reasons-responsive just in case it is regularly receptive, and overall reactive, to reasons for action. To illustrate, suppose Suzy wants to go to swim in the ocean and sets out to leave her house. As she's walking by the window, she realizes that a storm is coming. So, she changes her mind and stays at home. This suggests that Suzy's deliberative process—i.e., the 'mechanism' that issues her behavior—is receptive and reactive to reasons for action. When she has a desire to go to swim, absent any outweighing reasons, it recognizes this reason and issues a behavior accordingly with that

[62] Another leading compatibilist view is the sanity theory (Wolf 1987, 1990), which involves, in addition to the requirements of the deep self-view, an epistemic condition. Since I discussed the epistemic condition for responsibility in the previous section, I won't further discuss the sanity theory.

[63] See Frankfurt (1971), Mitchell-Yellin (2015), Shoemaker (2003), and Watson (1975) for various versions of the deep-self (or the hierarchical mesh) view.

reason. But upon realizing that there is an outweighing reason, it issues a different behavior accordingly with the outweighing reason.

Incompatibilist—i.e., libertarian—theories of control can be divided into three categories: event-causal theories, agent-causal theories, and noncausal theories.[64] According to event-causal theories, the control in question requires nondeterministic causation by apt mental states. That is, an agent controls her act just in case mental events that involve her reasons—i.e., her beliefs, desires, and intentions—nondeterministically cause her act. According to agent-causal theories, an agent controls her act just in case she nondeterministically causes her act. The agent is said to be an object rather than an event. Hence, although the causing of her act might involve antecedent mental events involving her reasons, the agent must be the ultimate cause of her own choice of action. According to noncausal theories, control doesn't (or need not) involve causation. It's argued that every action either is or begins with a mental action—i.e., choice. Making the choice doesn't involve exerting any causal power, and we make the choice ours simply by performing it, or by being the subject of the choice. And as long as this choice is undetermined by previous events, it's a freely made choice.

There's much more to be said about these theories but this suffices for our purposes. Notice that these theories flesh out responsibility relevant control ultimately in terms of factors internal to agents—beliefs, desires, cares, judgements, intentions, a specific relationship among

[64] For event-causal theories, see Balaguer (2004), Ekstrom (2019), and Kane (2016). For agent-causal theories, see Chisholm (1966), Clarke (1993), Griffith (2010). For noncausal theories, see Ginet (1997), Palmer (2016), Widerker (2018).

them or their causing one's choice of action in a certain way, one's capacity to be responsive to reasons, or one's causing one's choice of action or one's simply choosing in a certain way. But I need to address various potential objections before I can more firmly conclude that these theories don't require factors external to agents. Below I will discuss three objections.

The first objection concerns the guidance control theory. One might worry that reasons-responsiveness depends on one's response to moral reasons, and at least some moral reasons (e.g., special obligations, reasons to help others, reasons not to hurt others) are at least partly external to agents.[65] If this is correct, then the guidance control theory requires factors external to agents. However, a response to external moral reasons, or more broadly to what the external world is actually like, isn't necessary for moral responsibility. Consider an assassin firing her weapon to kill a victim. Unbeknownst to her, she was drugged half an hour ago. The only effect of the drug is that it made it seem to her falsely that the victim was standing right there as she fired her weapon. Intuitively, she is blameworthy despite the fact that her attempt to kill the victim was unsuccessful. But notice that she is not exactly responding to what the world is like, nor for that matter to moral reasons external to her. Consider also someone plugged to an experience machine. Imagine that for all he knows, he lives in a real world and everyone he interacts with are real people. In that simulated world, he is an assassin. He goes around 'killing' innocent people for money. Intuitively, he is blameworthy. We would hardly treat him any differently than any other real-world assassin if we unplugged him from the

[65] Cf. Fischer and Ravizza (1998:252-3). One might also object that the requirement for the reasons-responsive mechanism to be one's own also involves elements external to agents (cf. Fischer and Ravizza 1998:241-3). I'll discuss the so-called mechanism ownership condition below.

experience machine. But notice again that he is not responding to what the world is like. In fact, his perception is radically and systematically mistaken. He is *never* responding to what the external world is like. It follows that reasons-responsiveness cannot depend on responding to what the external world is like. Similarly with the conclusion in the previous section, what seems to matter ultimately is one's response to what one takes (or should have taken) the world to be like.

The second objection concerns an especially suspicious element above—i.e., (in)determinism. Compatibilists hold that the responsibility relevant control is compatible with determinism while incompatibilist hold that it's not. As it's typically understood in this debate, determinism is the idea that at any moment the state of world and the laws of nature entail a unique future, and indeterminism is the idea that determinism is false. Hence, (in)determinism is a feature not only of agents but of the world. It would then appear that something external to agents is a crucial factor for responsibility relevant control. However, this is mistaken—or so I will argue.

Let me first introduce some terminology. By "global (in)determinism," I will mean (in)determinism that obtains *only* outside an agent. By "agential (in)determinism" I will mean (in)determinism that obtains *only* within an agent. I will argue that global (in)determinism is irrelevant for responsibility relevant control. What matters is only agential (in)determinism, and hence the concern regarding (in)determinism in the debate over the control condition doesn't threaten responsibility internalism.

As it is typically understood in the contemporary literature, compatibilism is the idea that agents can be free and morally responsible even if determinism is true. That is, regardless of whether determinism obtains, we can still have the responsibility relevant control.[66] This suggests that (in)determinism is irrelevant to the compatibilist theories of control. Indeed, compatibilists sometimes mention this feature of their theories as an advantage over libertarian theories (Fischer 2007:47). It's also worth noting that a significant number of libertarian theorists hold that one or the other compatibilist theory mentioned above identify a *necessary* condition for responsibility relevant control.[67] But this is viable only if compatibilism is consistent with indeterminism. Hence, the compatibilist theories above are consistent with (in)determinism.

Moreover, even if determinism is not just consistent with but necessary for control, it's unclear why the required determinism should be global determinism. Even if determinism is necessary for control, it is hard to think that some undetermined particles somewhere far in the universe could undermine this control.[68] If this is correct, then it is also hard to think that some undetermined particles right outside the agent and about to move into the agent could

---

[66] Some classical compatibilists thought that determinism is not only compatible but also necessary for free will. Starting from around 1960s, more and more compatibilists gave up the idea that determinism is necessary. An influential factor in this turn of events has been quantum mechanics and indeterministic interpretations of physics becoming more and more popular. However, even for those who in the past defended that determinism is necessary for free will, at least for some of them, it's unclear whether they required global determinism or agential determinism. For instance, it's often taken that Hobart (1934) argues that determinism is necessary for free will. But, as Cyr (forthcoming)—himself a compatibilist—argues, it's best to interpret Hobart as requiring not (what I called) global determinism, but agential determinism. The idea that freedom is compatible with both determinism and indeterminism is sometimes called 'supercompatibilism' (Vargas, 2012, p.420 and fn.4.) which is the majority position among contemporary compatibilists.

[67] See O'Connor and Franklin (2021), section 2.5, for a discussion on this and a survey of libertarian theorists who build their views on one or another compatibilist theory.

[68] van Inwagen (1983:126) makes a similar point for the relevance of determinism for incompatibilist theories. He says, "if determinism is incompatible with free will, so is the thesis that everything except one distant particle of matter is determined."

undermine her control. Mere distance (spatial or temporal) of these particles, as long as they're outside the agent, doesn't seem to be relevant to agential control. Whether these particles are determined or undetermined, they're completely outside the agent's control, and as such, their moving into the agent one way or the other couldn't make a difference to the control she exercises afterwards.

To illustrate, and to bolster the point here, consider a case of instant creation. Imagine that an agent is instantly created and, immediately after creation, she finds herself in an environment where she must make a choice. The particles in the environment could be determined or undetermined. The only difference between these two possibilities is that while the deterministic scenario ensures that there's a single way for particles to move into the agent, in the indeterministic scenario there are multiple options. But this doesn't seem to be relevant to the control she will exercise afterwards. Notice that we're not imagining a unique environment in which she finds herself. We could have easily imagined a totally different environment. In this new environment, if determinism obtains, there's again a single way for the particles to move into the agent. But the way they move in this new scenario is different than the way the move in the previous deterministic scenario. It's hard to think that the agent exercises control over her choice in one of these deterministic scenarios but not the other. Hence, the exact way the particles move into the agent is irrelevant to the control she exercises. They could move one way or the other. And this is just what happens when the particles in the environment are undetermined—they could move one way or the other. It follows that even if some form of determinism is necessary for control, it's not the global determinism that's necessary.

Turning now to incompatibilist theories of control—i.e., libertarians—, they require indeterminism for responsibility relevant control. However, global determinism is consistent with all the libertarian theories above. While some event-causal theories require that determinism take place in one's *deliberation*, others require that indeterminism take place *after deliberation and before the choice* of action is caused. It is sufficient for agent-causal theories that an agent's causing of her choice of action be uncaused or not be deterministically caused by prior *mental events*. It is sufficient for noncausal theories that an agent's simply choosing—instead of *causing* her choice—be undetermined by prior *mental events*.

Moreover, similarly as above, even if indeterminism is necessary for control, it's unclear why the required indeterminism should be global indeterminism. One major libertarian concern about determinism is that an entirely deterministic world doesn't leave room for agents genuinely contributing to their own choices of action. This is because, it's argued, such a universe dictates choices of action before agents even begin their deliberation. But even if global determinism obtains, as long as agential determinism doesn't, what comes before one's choice or deliberation doesn't determine—i.e., doesn't 'dictate'—how one deliberates or chooses. This suggests that even if indeterminism is necessary for responsibility relevant control, it's not the global indeterminism that's necessary.

An objector might rightly point out that the libertarian worry about determinism isn't only a *backward-looking* one—i.e., prior to and including deliberation or choice. The other concern they have about determinism is a *forward-looking* one—i.e., that determinism doesn't leave room for alternate possibilities. And according to the Principle of Alternate Possibilities (**PAP**), availability of alternate courses of action is a necessary condition for moral

responsibility. This seems to imply morally responsibility requires some elements external to agents—i.e., multiple courses of future. However, recall that many libertarians deny **PAP**. This suggests that rejecting **PAP** is consistent with libertarian theories of control. Moreover, as I argued in the previous chapter, since extramental elements in one's causal responsibility are irrelevant for moral responsibility, alternate courses of future are also irrelevant for moral responsibility. Hence, **PAP** is false.

It seems then that global (in)determinism is irrelevant for both compatibilism and incompatibilism about responsibility relevant control.

The third objection stems from one among the most pressing arguments against various (especially the compatibilist) accounts of control—i.e., the ones based on manipulation cases. The relevance of this for our purposes is that, in response, some philosophers suggest that an appropriate account of control must involve some historical elements, or elements external to agents. The so-called manipulation argument comes in different forms. I'll discuss one that's due to Mele.[69] Consider the following case.

> **(Brainwashed Beth)** Beth has an exceptionally good moral character. She leads a morally good life—she never means harm to anyone, she's very kind and always extends herself to help others. One night while she's asleep, a team of psychologists implant new beliefs, desires, and values in Beth after erasing hers—i.e., they brainwash her. They leave Beth's memory intact, and we can assume that Beth remains moderately reasons-responsive as before. Next morning, Beth wakes up with a strong desire to kill her neighbor George. Killing George is also what she wants deep down and is consistent with her new set of beliefs and values. So, she goes ahead and kills George. The following night the psychologists reverse the brainwashing. Beth wakes up in the morning having the good moral character that she had before.

[69] See Mele (2009, 2016, 2019:ch. 4). Another well-known manipulation argument is Pereboom's Four-Case Argument (Pereboom 1995, 2014:ch.4).

Notice that when she killed George, Beth satisfied the conditions for the compatibilist theories of control mentioned above. But intuitively, it's argued, she's not morally responsible for killing George. Compatibilist replies to this worry fall into two categories—the hard-line reply and the soft-line reply.[70] While the hard-liners hold that Beth did have the responsibility relevant control when she killed George and she is blameworthy, soft-liners disagree.

But if Beth doesn't have the responsibility relevant control despite the fact that she meets all compatibilist conditions internal to agents, there must be some further—i.e., external—condition(s) for compatibilist theories of control, or so the soft-liners reply. For instance, driven mainly by this concern, some soft-liners argue that the guidance control theory involves a mechanism ownership condition. That is, responsibility relevant control requires not only that the mechanism that issues one's behavior be moderately reasons-responsive but also that the mechanism be one's own. It's argued that one makes a mechanism their own by *taking* responsibility for the behavior that the mechanism issues. One can achieve this over time, and without reflection (e.g., as a result of one's upbringing and moral education) or via reflection (on one's choices and actions, and others' attitudes towards one's choices and actions). And

[70] This terminology is due to McKenna (2008) who takes the hard-line approach. Among the other hard-liners are Cyr (2019), Frankfurt (2002), Khoury (2014), Tierney (2013), and Watson (1999). Among the soft-liners are Demetriou (2010), Fischer and Ravizza (1998), and Mele (2019) (though Mele is agnostic about compatibilism (cf. Mele 2019:3)). The debate between hard-liners and soft-liners is basically a debate between externalists and internalists. In the literature, externalists are sometimes called "historicists" because they argue that certain historical conditions affect the control condition (or moral responsibility). Internalists are sometimes called "snapshot theorists" or "time-slice theorists" because they argue that all that's relevant to the control condition (or moral responsibility) is an agent's snapshot or time-slice properties. I think this a bit misleading. One can have external properties at a given time-slice. One can also have historical but internal properties. For instance, given a four-dimensionalist space-time worm picture of an agent, one's past mental states can be both historical and internal to the agent. So, I suspect some so-called historicists might properly be called internalists, or that there might be internalist theories that involve historical elements. But it's rather clear in the literature that many so-called historicists require elements that are genuinely external to agents. One last note: the debate between the historicists and snapshot theorists is *not* a debate merely among compatibilists, and is independent of the debate between compatibilists and incompatibilists. But for simplicity, I write as if it's a debate among compatibilists.

since the mechanism that issued Beth's behavior—i.e., killing George—is *not* one that Beth takes responsibility for, Beth lacked responsibility relevant control when she killed George and hence isn't blameworthy.

However, this reply doesn't adequately address the worry raised by manipulation cases but merely pushes the problem back. Here is why. The underlying worry in **(Brainwashed Beth)** is that agents other than Beth are in charge of what she does. And notice that the temporal distance of these other agents—i.e., the manipulators—to the manipulatee is irrelevant. If the manipulators set up a system a few days before that fateful night and ensured that it would kick in to do the brainwashing the night that the original brainwashing occurred, the worry would remain that Beth isn't morally responsible for killing George—*if* she's in fact not morally responsible. In that case, consider this:

> **(Seth-to-Kill)** Seth was created by a super-powerful being in a deterministic world such that when Seth turns thirty, he will kill Jerry. It is also determined by this super-powerful being that the mechanism that issues Seth's behavior is moderately reasons-responsive, and that earlier in his life Seth will go through a process after which he will take responsibility for this mechanism. Years later, Seth turns thirty and kills Jerry.

Notice that just like in **(Brainwashed Beth)**, an agent other than Seth is in charge of what he does. And if the temporal distance between the manipulator and the manipulatee is irrelevant, **(Seth-to-Kill)** is the same as **(Brainwashed Beth)** in all relevant respects. It follows that if Seth is blameworthy, then Beth is blameworthy all the same, and if Seth is not blameworthy, neither is Beth.[71] But then it is unclear what the externalist requirement in question brings to the table in the face of the underlying worry in manipulation cases. This is because while **(Brainwashed**

[71] This is similar to the worry sometimes raised by the so-called "zygote argument." See Mele (2016:71–2, 2019: 83–4).

Beth) and **(Seth-to-Kill)** are relevantly similar in all respects (i.e., if the former is a manipulation case, so is the latter), and Beth doesn't meet the above externalist requirement and Seth does, Beth and Seth are on a par regarding their moral responsibility.[72]

Notice also that this is not a problem only for the above suggested externalist requirement but generalizes over all potential externalist requirements. As Watson (1999:360-1) puts:

> [C]onsider any compatibilist account of the conditions of free agency, C. It is possible for C to obtain in a causally deterministic world. If that is possible, then it is possible that a super-powerful being intentionally creates a C-world, by bringing about the relevant antecedent conditions in accordance with the relevant laws.

So, for any potential compatibilist control condition—even those that are external to agents—that condition can be brought about by someone other than the agent herself. It follows that externalist requirements can hardly add anything to the compatibilist reply against the underlying worry in manipulation cases—which is basically what motivates such requirements. If this is correct, compatibilists should reject externalist requirements.

Moreover, internalist compatibilists still have much to say about manipulation cases. For instance, consider **(Brainwashed Beth)** again. Let's call Beth pre- and after-manipulation "$Beth_1$," and the manipulated Beth "$Beth_2$." As Khoury (2013:745-51) points out, much of the intuitiveness of the idea that Beth isn't morally responsible for killing George comes from the fact that $Beth_1$ is a morally decent person. Even if $Beth_1$ and $Beth_2$ are the same person, the two are total opposites of one another concerning their moral characters. $Beth_1$ is the furthest thing

[72] Notice that I don't necessarily deny the mechanism ownership condition. I deny it if, or to the extent that, it requires elements external to agents.

from a murderous character and certainly doesn't seem to be morally responsible for killing George. But this is consistent with compatibilism. Compatibilists are not committed to thinking that one always remains blameworthy for one's past actions.[73] Consider the fact that after people genuinely repent, or change their morally undesirable character, we often think that they are no longer blameworthy for what they might have done in the past. Plausibly, a good explanation for this is that once people genuinely repent or change their ways, they don't share much with the morally undesirable character that they might have had in the past. If this is on the right track, notice that $Beth_1$, concerning her moral character, doesn't share anything with $Beth_2$. It follows that we shouldn't think of $Beth_1$ as blameworthy for killing George, and this is consistent with compatibilism.

Indeed, the suggestion above could only be the beginning and part of what internalist compatibilists could say about manipulation cases. This is because there's still the question as to whether $Beth_2$ is blameworthy. But whatever the appropriate reply here might be, the externalist position doesn't seem to have much to contribute to it, and this suffices for my purposes.

To sum it up, in this section, I presented the leading theories of responsibility relevant control and argued that they do not require anything external to agents. I considered three objections. One, reasons-responsiveness depends on one's response to what the world is

[73] Notice that the idea that one always remains blameworthy for one's past actions assumes a relationship between personal identity and (diachronic) moral responsibility. But it's dubious that there's a metaphysically significant relationship between the two. See Shoemaker (2012) for an argument that responsibility does not presuppose personal identity. See Khoury and Matheson (2018) for an argument that personal identity isn't sufficient for responsibility. As they argue, it's false that 'once blameworthy always blameworthy.'

actually like. Two, (in)determinism is relevant to control. Three, arguments from manipulation cases motivate external conditions for control. However, I argued that first, responding to what the world is actually like isn't necessary for moral responsibility. Second, (in)determinism that obtains outside the agents is irrelevant to control. Third, an external condition is no help against arguments from manipulation cases. It follows that these theories of control, in their most plausible forms, don't require anything external to agents. Hence, I conclude that responsibility relevant control depends only on factors internal to agents.

3. Conclusion

In this chapter, I discussed two of the conditions for moral responsibility: the epistemic condition and the control condition. I argued that neither of them requires anything external to agents. Hence, both the epistemic condition and the control condition depend only on factors internal to agents.

Chapter 6: Responsibility Gap: Not New, Inevitable, Unproblematic

Abstract:
Who is responsible for a harm caused by AI, or a machine or system that relies on artificial intelligence? Given that current AI is neither conscious nor sentient, it's unclear that AI itself is responsible for it. But given that AI acts independently of its developer or user, it's also unclear that the developer or user is responsible for the harm. This gives rise to the so-called *responsibility gap*: cases where AI causes a harm but no one's responsible for it. Two central questions in the literature are whether responsibility gap exists, and if yes, whether it's morally problematic in a way that counts against developing or using AI. While some authors argue that responsibility gap exists, and it's morally problematic, some argue that it doesn't exist. In this paper, I defend a novel position. First, I argue that current AI doesn't generate a new kind of concern about responsibility that the older technologies don't. Then, I argue that responsibility gap exists but it's unproblematic.

Consider self-driving cars, lethal autonomous weapons, candidate screening tools, medical systems that diagnose cancer, and automated content moderators. These new technologies go beyond merely executing certain commands. They learn, make decisions, and act on their own. While some people celebrate these new technologies, others find them somewhat unwelcome. The idea of a 'killer robot' or a 'robot' deciding whether you get hired can be rather chilling.

Let's call these autonomous systems or machines that rely on artificial intelligence "AI." One among the moral questions that come along with AI concerns responsibility. Imagine that an autonomous weapon kills a non-combatant, or a self-driving car injures a pedestrian.[74] Who's responsible for these harms? Current AI is far from being conscious, sentient, or

[74] On March 18, 2018, a pedestrian, Elaine Herzberg, was killed by an Uber self-driving car in Tempe, Arizona. The incident is the first recorded case of a pedestrian fatality associated with a self-driving car. (Wakabayashi 2018.)

possessing agency similar to that possessed by ordinary adult humans. So, it's unclear that AI is responsible for a harm it causes. But given that AI gathers new information and acts autonomously or unpredictably, it's also unclear that those who develop or deploy AI are responsible for what AI does. This leads to the so-called responsibility gap (**RG**): that is, roughly, cases where AI causes a harm, but no one is responsible for it.

Two central questions in the literature are **(Q1)** whether **RG** exists and **(Q2)** whether **RG** is morally problematic. Some authors hold that there is no **RG** (or it's dubious that **RG** exists) and hence that there is no **RG**-related worry against developing or using AI.[75] And some authors hold that **RG** exists, and it counts against developing or using AI.[76] For instance, imagine again a self-driving car injuring a pedestrian. If **RG** exists—i.e., if no one is responsible for the harm caused by AI—then one might worry that we can't justly hold anyone responsible for what happened to the pedestrian. But it seems odd that there's no one to be justly held responsible in such a case. Consider also that a just war may require holding those who kill non-combatants responsible. But if no one is responsible for it when an autonomous weapon kills a non-combatant, fighting a war that involves use of AI risks being unjust (Sparrow 2007).

The above discussion illustrates two of the possible positions defended in the literature: **(1) RG** exists, and it's morally problematic, and **(2) RG** doesn't exist, and hence there's no **RG**-related problem for AI.[77] In this chapter, I defend a novel position: **(3) RG** exists, but it's

---

[75] Cf. Simpson & Müller (2016), Burri (2017), Köhler, Roughley & Sauer (2017), Himmelreich (2019), Tigard (2021), Himmelreich & Köhler (2022), and Königs (forthcoming).

[76] Cf. Matthias (2004), Sparrow (2007), Roff (2013), Danaher (2016). The name "responsibility gap" is due to Matthias (2004). Danaher (2016) argues for "retribution gap" which is supposed to be distinct from **RG**. However, it's unclear that retribution gap is fundamentally different from, or doesn't presuppose, **RG** (cf. especially p.306).

[77] I don't mean that everyone in the literature falls neatly into the category of those that defend either **(1)** or **(2)**. Notice that in addition to the question of whether **RG** exists, there's also the question of how often it occurs *if* it

*un*problematic. Here, more specifically, are the three theses I defend below. One, current AI doesn't generate a kind of concern about responsibility that the older technologies don't (**§1**). Two, a harm caused by AI never *affects* anyone's responsibility—i.e., it doesn't change anyone's *degree* of responsibility, nor is it what anyone is responsible *for* **(§2)**. Three, this isn't worrisome for developing or using AI. On the contrary, holding otherwise commits one to various implausible views **(§3)**.

Here are a few clarifications before I begin. First, AI in question is the current AI or AI that's possible in the near future, not *strong* AI (or super-intelligent AI) that might be developed in distant future. Second, I won't argue that developing or using AI isn't morally worrisome. All I'll argue is that **RG** doesn't generate such a worry.[78] Third, the relevant sense of moral responsibility in **RG**, and what I'll henceforth refer to by "responsibility" unless I indicate otherwise, is basic desert responsibility (or accountability).[79]

---

exists. Hence, one might instead hold **(2*)** that **RG** exists but occurs too rarely, hence it's not worrisome, or **(1*)** that **RG** occurs not whenever AI causes a harm but often enough, hence it's worrisome. It's not always clear whether the thesis defended is **(2)** or **(2*)**, or **(1)** or **(1*)**. This is partly because in some cases the line between the author's argument for or against the existence of **RG** versus their argument for or against whether **RG** is problematic isn't drawn very clearly. At any rate, some 'deniers' of **RG** would defend something similar to **(2*)** (e.g., Simpson & Müller (2016), Nyholm (2018a), and Königs (forthcoming)). And some 'accepters' of **RG** would defend something similar to **(1*)** (e.g., Sparrow 2007:69-70). Also, some authors assume that **RG** exists, but deny that it's problematic (cf., Goetze 2022, and Kiener 2022). None of this affects the novelty of the position I defend in this paper.

[78] Cf. Sparrow (2016) for a critical discussion of the extant moral arguments, and Sparrow's novel argument, against lethal autonomous weapons (LAWs) in particular. Cf. Jenkins & Purves (2016) for a response to Sparrow. Cf. Simpson & Müller (2016), and Müller (2016) for arguments in favor of LAWs, which also involve discussions of **RG**. Cf. Müller (2021) for a survey of moral concerns that arise in the context of AI.

[79] Some authors employ a combination of accountability, attributability, and answerability conceptions of responsibility while discussing **RG** (cf. Himmelreich 2019, Tigard 2021). However, many if not most authors from various sides of the debate understand **RG** to be about accountability (cf. Sparrow (2007:71), Roff (2013:353-4), Müller (2016:73), Burri (2017:176-6), Köhler, Roughley & Sauer (2017:52-3), Goetze (2022), and Königs (forthcoming)). Moreover, whether **RG** exists in this sense and, if yes, whether it counts against developing or using AI are interesting questions on their own.

1- Responsibility: AI versus Older Technologies

Consider older technologies such as cars, clocks, calculators, rifles, catapults, wheelbarrows, and shovels. Certainly, AI is different from these technologies in crucial respects. But is it so different as to generate a new kind of concern about responsibility that the older technologies don't?

To answer this question, we need to look at AI's feature that's relevant for responsibility, which—as it's typically emphasized—is its autonomy: once deployed, AI processes new information and acts on its own based on this new information. AI's autonomy could be relevant to AI's own responsibility, or to others' responsibility. So, first, one might think that since responsibility requires autonomy, AI's autonomy might mean that AI itself is a kind of being that's responsible for its actions. However, it's relatively uncontroversial that the current AI isn't this sort of being.[80] Even if AI acts autonomously in some sense of the word, its autonomy isn't robust enough to ground responsibility. Consider, for instance, that animals and children act autonomously in much the same way as AI—i.e., they learn and act on their own. But neither animals nor children are responsible for their actions. Hence, AI, regarding whether it's a responsible being, is no different than the older technologies.

Second, AI's autonomy can be relevant to others' (e.g., developers', users') responsibility. After all, AI is a tool, and developing or using a tool can be relevant to one's responsibility. But how exactly is this different than developing or using an older technology?

---

[80] Cf., e.g., Roff (2013:353-4), Danaher (2016:304), Müller (2016:73), Nyholm (2018a:1209), Himmelreich & Köhler (2022), Königs (forthcoming). Cf. Sars (2022) for an argument that AI that can be held responsible can be designed in future.

It's tempting to respond that AI's autonomy implies that developing or using AI can cause unpredictable outcomes. But developing or using any old technology can also cause unpredictable outcomes. Despite all precautions, a malfunction in a car can lead to an accident, an arrow might hit a non-combatant, and a broken clock can make you late for work.

One might object that AI's unpredictability is feature, not a bug, whereas, say, a malfunction in a car is a bug. That is, an autonomous tool is an inherently unpredictable tool. But it's unclear why a tool's being inherently or non-inherently unpredictable should make a difference to developers' or users' responsibility. One might respond that it's more difficult to predict what an inherently unpredictable tool will cause. However, first, if this is meant to be an empirical claim, it's unclear that AI is actually more unpredictable. And even if AI is more unpredictable, the difference between how AI is relevant to others' responsibility versus how the older technologies are relevant to others' responsibility would be a matter of degree and not of kind. Consider also that the older technologies are not all equally predictable either. Second, a poorly designed AI certainly might cause many unpredictable outcomes. But this is true of any poorly designed technology—especially considering not only the immediate consequences of using a technology but the consequences in the long run. And if what's in question is a well-designed AI, it's dubious that using it causes significantly more unpredictable outcomes than using an older technology. In fact, using a well-designed AI might predictably cause even better outcomes than using an older technology. For instance, it's likely that self-

driving cars will decrease the number of car accidents, and autonomous weapons will decrease the human cost of war.[81]

It seems then that AI doesn't generate a kind of concern for others' responsibility that the older technologies don't generate for others' responsibility. Hence, given also that AI and the older technologies are on a par regarding their own responsibility, we can conclude that AI doesn't generate a kind of concern about responsibility that the older technologies don't. Here are two of the significant implications of this conclusion. One, at least a big part of our initial worries about responsibility in the context of AI is likely unfounded. Two, it's unlikely that the underlying concern in responsibility gap **(RG)** is something uniquely AI-related. And hence, even if **RG** exists, it's unlikely that our existing philosophical tools are inadequate to render it unproblematic. Both these points will be further evident below. Next, I turn to my argument that **RG** exists.

2- Responsibility Gap: Inevitable and Ubiquitous

Here's a more precise characterization of **RG**:

> Responsibility gap occurs when someone develops or deploys AI which then causes something morally unwanted, X, but no one's responsibility is affected by X.

By "no one's responsibility is affected by X," I mean no one is *more* or *less* responsible because of X, nor is X what anyone is responsible *for*.[82]

---

[81] Cf. Arkin (2010), Jenkins & Purves (2016), Müller (2016), Burri (2017), Nyholm (2018b), and Himmelreich & Köhler (2022) for further discussion.

[82] Most authors characterize **RG** solely in terms of what one is responsible *for* and leave out the *degree* of responsibility. (To my knowledge, the only exception is Köhler, Roughley & Sauer (2017:54).) My characterization of **RG** is otherwise in line with other characterizations. Cf., e.g., Matthias (2004, p.175, p.177), Burri (2017:175-6), Himmelreich (2019, p.731, p.734), Kiener (2022), Königs (forthcoming).

In previous chapters, I argued that the actual consequences of actions (or causal responsibility) are irrelevant for responsibility—that they never affect the degree to which we are blameworthy, nor are they what we are blameworthy for. Given that "X," in the characterization of **RG** above, refers to the actual consequences of one's actions, **RG** occurs whenever AI causes anything. That is, **RG** is inevitable and ubiquitous.[83]

Here's also a further argument that actual consequences of an action are *not* relevant for responsibility. Suppose the actual consequences of an action *are* relevant for responsibility. The question is: Which actual consequences specifically are relevant? First, the idea might be that: **(AC)** *all actual consequences* of an action are relevant for responsibility. But **AC** seems false. Consider that totally random consequences can result from an action—not to mention the numerous consequences that follow further (e.g., many years later) down the causal chain. Suppose you touch the switch to turn on the lights in your room, but unbeknownst to you, the switch is connected to a torture machine in the next room, which results in an innocent victim's being tortured. It's implausible that the victim's being tortured makes you blameworthy or increases your blameworthiness.

[83] Köhler, Roughley & Sauer (2017:61-2) reject an argument that **RG** exists on grounds that there's resultant moral luck (**RML**) which they assert without argument. (Curiously, they also note that there's "something unjust" in **RML**, but they deny dealing with it claiming that the question goes beyond the scope of their paper (p.62).) However, as I'll argue below, **RG** remains inevitable even assuming **RML**. Hevelke & Nida-Rümelin (2015) don't explicitly discuss **RG**. But they argue that **(i)** the owner of, or a passenger in, a self-driving car isn't the (only) one who has a duty in reparations in case of an accident caused by the car because **(ii)** there is no **RML** (pp.626-8). However, although **(i)** might be the appropriate policy to adopt, **(ii)** doesn't entail **(i)**. Also, they reject **RML** arguing that **RML** is, in effect, committed to the following claim: If unwanted consequences follow from an action that's all-things-considered right but against which there is *even the smallest reason*, these consequences increase one's blameworthiness. But I doubt that proponents of **RML** would accept this. At any rate, **RML** is consistent with the idea that one isn't blameworthy unless one's (attempted) action is all-things-considered wrong (or suberogatory). (Cf. Hartman (2017, p.34, pp.91-2), a proponent of **RML**, on this.)

Second, the idea might be that: **(EC)** only the *expected consequences* are relevant for responsibility—whether it be the consequences expected by the agent or the consequences that the agent is reasonably expected to expect. And to my mind, **EC** is also the most plausible way to avoid the above objection to **AC**: Since it's neither the case that you expect, nor is it reasonable to expect that you expect, that touching the switch will result in an innocent person's being tortured, you're not blameworthy. However, the problem is that if **EC** is true, then it's false that the actual consequences are relevant for responsibility. This is because expected consequences are not actual consequences. One's beliefs about the potential consequences of an action can be false. The beliefs that are reasonable to expect one to have regarding the consequences of one's action can also be false. By analogy, consider that it might be reasonable to expect that many people have certain current scientific beliefs, while they actually don't, but nonetheless these scientific beliefs can be false.

Third, the idea might be that: **(AEC)** only *the actual* (or *obtaining*) *consequences among the expected consequences* are relevant for responsibility. Notice that **AEC** draws a line between the actual consequences that *are*, and those that are *not*, among the expected consequences. The question is 'why.' Why should only the expected consequences among the actual consequences matter for responsibility? First, the most straightforward answer is that it's because they are expected. But then what ultimately matters is the expectancy, and not the actuality, of the consequences. Thus, **AEC** collapses back to **EC**. Second, one might think that they matter because they are actual consequences. But, as we saw with **AC** above, *mere* actuality of the consequences can't matter for responsibility. Thus, **AEC** turns out false just like **AC**. Third, one might think that they matter because they are expected and actual. But this is

mere insistence that actuality of the consequences somehow matters for responsibility. Given also the reasons in previous chapters, we have good reasons to reject this insistence. I submit again then that the actual consequences of an action are irrelevant to one's degree of responsibility and what one is responsible for.

Here's one further thing to note. Suppose the actual consequences of an action are relevant for responsibility. Can we avoid **RG**? We can, but only if we accept **AC**—that all actual consequences are relevant for responsibility. However, then we'd be committed to a highly implausible claim. One might instead accept **AEC**—that only the actual consequences among the expected consequences are relevant for responsibility. Although as I argued above that **AEC** ultimately fails, I suspect **AEC** is more palatable than **AC**. But even given **AEC**, we can't avoid **RG**. This is because it's always possible that AI brings about consequences that were *not* among the expected consequences. Indeed, there's still the question regarding how often it will be the case that AI brings about unexpected consequences. And the answer might be that it's rare. One might then argue that **RG** isn't problematic *partly* because it isn't ubiquitous. Be that as it may, **RG** itself remains inevitable.[84] In the next section I'll assume, as argued above, that **RG** is both inevitable and ubiquitous, and argue that this isn't problematic.

3- Responsibility Gap Is Not Problematic

[84] The foregoing discussion also addresses two of the following questions in the literature. One, Himmelreich (2019:739) suggests that proponents of resultant moral luck (**RML**) might avoid **RG**. Proponents of **RML** typically don't take on the question of which actual consequences specifically *do* affect responsibility. But if they accept **AC**, it makes their view even more implausible, and I'm not aware of any proponent of **RML** openly accepting **AC**. At least one proponent of **RML** openly accepts something like **AEC** (Hartman, 2017:92-3). And as argued above, even given **AEC**, **RG** remains inevitable. Two, those who accept the existence of **RG** typically don't take on the question of when exactly **RG** occurs. Königs (forthcoming) challenges accepters of **RG** on this. If I'm right, the answer is always—**RG** occurs whenever AI developed or deployed by someone causes any outcome. If **AEC** is true, **RG** occurs whenever AI causes something unexpected.

Consider the following case:

> **(Drone)** A soldier (let's call her "Soldier") deploys a drone which flies out to the warzone. A non-combatant (let's call him "Victim"), of his own will, is in the nearby area. The drone falsely identifies him as threat, and kills him.

Notice that **(Drone)** is relevantly similar to other AI cases such as when a self-driving car carrying a passenger or its owner gets into an accident, or when a candidate screening tool eliminates the perfect candidate for the job. Hence, to keep the discussion clear and concise I'll focus on **(Drone)**, but what I'll say about **(Drone)** applies equally to all other relevant cases.[85]

Notice also that **(Drone)** may or may not involve a 'foul play.' So, first, assume that there's no foul play involved. That is, Soldier made sure that the drone passed all the inspections, she didn't intentionally omit any of her duties, she didn't expect Victim's death, nor was it reasonable to expect that she expect Victim's death, and so on. Everyone should agree that neither Soldier nor anyone else is responsible for Victim's death. Hence, insofar as responsibility gap **(RG)** occurs when AI deployed by someone causes a harm, but no one is responsible for it, **(Drone)** *is* a case of **RG**. But everyone should also agree that this isn't problematic. If there's no foul play involved, **(Drone)** is no different than any other unfortunate case where things just go wrong despite all precautionary measures. Though even if one disagrees, this isn't troubling for me. Cases like **(Drone)** that *do* involve a foul play are the

[85] Notice that **(Drone)** involves a *user* of AI, and not a *developer*. But nothing of relevance changes if Soldier is the developer *and* the user. Also notice that **(Drone)** involves only one person (as user or developer), and not a group of people. Although things would be more complicated if **(Drone)** involved a collection of people, the complications would *not* be due to AI but due to questions about collective responsibility. (Cf. Khoury (2017) and Smiley (2022) for discussions on collective responsibility.) Hence, I leave such cases aside though what I say below is helpful in such cases too. Also, some suggest that in such cases **(P)** one is responsible to the extent that one is a cause of the harm (cf. Köhler, Roughley & Sauer, 2017:58-9). Notice that **P** just is **Proportionality** which, I argued, is false.

seemingly more problematic and interesting cases, and what I say about the latter largely applies to the former kind of cases.

Second, assume that **(Drone)** involves foul play. That is, Soldier didn't ensure that the drone passed all the inspections, intentionally omitted a duty, or foresaw Victim's death and believed correctly that it would be wrong to kill him, etc. but she deployed the drone anyway. Since I argued that **RG** occurs whenever AI causes a harm, I'm committed to holding that Soldier isn't responsible for Victim's death. And here are basically the two reasons why one might think this is morally problematic:

> **(MP1)** If Soldier isn't responsible for Victim's death, Soldier can't have any duties (e.g., to make amends with Victim's family). But it's unacceptable that no duties incur for Soldier after what's she's done.
>
> **(MP2)** If Soldier isn't responsible for Victim's death, we can't punish her (or otherwise hold her accountable) once she kills Victim in a way that involves foul play. And since **(Drone)** involves a war, there's another worry: A just war requires holding those who kill non-combatants responsible. Hence, if we can't hold Soldier responsible, we violate the norms of just war, which in turn implies that we waged an unjust war.[86]

In the rest of this section, I'll argue that **MP1** and **MP2** are misguided. Hence, **RG** isn't worrisome in a way that counts against developing or using AI. Let's begin with **MP1**.

MP1 asserts that unless Soldier is blameworthy for Victim's death, she can't have a duty regarding his death. The underlying assumption here is that being blameworthy for something is necessary for having a duty concerning that thing. But recall from **chapter 4** that this is false. To briefly reiterate, first, consider moral duties like telling the truth, keeping one's promises,

[86] Cf. Himmelreich & Köhler (2022, §2.1) and Königs (forthcoming) for other surveys of reasons found in the literature as to why **RG** might be problematic.

being non-maleficent, or helping those in need. None of them requires being blameworthy for anything. Second, consider that *mere* causal responsibility is often sufficient to generate a duty. Suppose you take someone else's property not knowing that it belongs to someone else. When it turns out later that it does belong to someone else, you have a duty to give it back, which doesn't require you to be blameworthy for taking it in the first place. Third, consider cases where, *through no fault of your own*, your teenage kid, pet, or farm animal causes harm to others. You're not blameworthy for the harm, but it's perfectly plausible that it's your duty to compensate for it. So, being blameworthy for something is *not* necessary for having a duty concerning that thing. Indeed, holding otherwise commits one to implausible claims.

Of course, it doesn't follow that Soldier doesn't have a moral duty after killing Victim. But it follows that *if* Soldier has a duty, it's dubious that this is because she's blameworthy *for* killing Victim. And recall again from **chapter 4** the two alternative suggestions about what might give rise to Soldier's duty. One, Soldier's duty arises at least partly from the fact that she's blameworthy for willing or acting to deploy the drone while, say, knowing that it will wrongly kill a non-combatant *plus* she's causally responsible for killing a non-combatant (Khoury 2018). Two, it could be that Soldier's duty arises from what Susan Wolf (2001) calls the '*nameless* virtue'—the virtue of *taking* responsibility for the consequences of one's actions *even if* those consequences are unforeseen, unintended, or somehow out of one's control. Both these suggestions are plausible and consistent with the existence of **RG**. It follows that **RG** is not an

impediment to potential duties incurring for Soldier.[87] Hence, **MP1** fails to show that **RG** is worrisome.

Let's now discuss **MP2**. Recall that the worry is that if Soldier isn't responsible for killing Victim, we cannot punish Soldier. And since **(Drone)** involves using AI in the context of war, there's an additional worry. If there's no one to hold responsible for killing Victim, Soldier's army waged an unjust war.

However, my argument for the existence of **RG** doesn't entail that we cannot punish Soldier. On the contrary, if I'm right, *that* Soldier is blameworthy, and the *degree* of her blameworthiness are fixed the moment she acts to deploy AI. What follows from my argument is that neither the fact that she's blameworthy nor her degree of blameworthiness changes *even if* Victim doesn't die. Hence, the reasons for punishing her or holding her accountable are in place regardless of what happens to Victim. Sure, we may lack firm reasons to believe that Soldier is blameworthy if Victim doesn't die. But this is a problem independent of **RG** or my argument for its existence. We might lack firm reasons to believe that someone is blameworthy even when their acts result in unwanted consequences. This is because, even if actual consequences are relevant for responsibility, no one holds that they are sufficient for responsibility. We need to know whether the person acted freely or satisfied certain epistemic conditions, and it's not always easy to find reasons to form strong beliefs about these

[87] I don't mean to imply that in real life cases the person who develops or deploys AI is the one and only person for whom a duty to compensate incurs. Depending on further details, there may be no duty that incurs, or the duty might incur for the group (e.g., army, company, or country) to which the person in question belongs. My claim is the more general one that **RG** itself isn't an impediment to incurring of such duties.

requirements that are wholly internal to agents. And they are wholly internal to agents because responsibility internalism, as I argued, is true.

So, we *can* punish Soldier or hold her accountable. Hence, the worry that **RG** leads to unjust war dissipates as well.[88] Here are two final things to note. One, *if* **RG** leads to unjust war, and hence AI shouldn't be used in war, then no older technology (e.g., cars, chariots, swords, rifles, spears, arrows, catapults) should be used in war. This is because, as argued above, AI doesn't generate a kind of concern about responsibility that the older technologies don't. To further illustrate, recall the above characterization of **RG**: Responsibility gap occurs when someone develops or deploys AI which then causes something morally unwanted, X, but no one's responsibility is affected by X. Suppose in this characterization we replace "AI" with an older technology and call what we characterize "**RG***." Does **RG*** exist? It sure does. Suppose instead of deploying AI, Soldier shoots an arrow into the warzone and Victim dies as a result. If Soldier took all the reasonable precautions, she's not responsible for Victim's death. And insofar as **RG*** occurs when using an old technology causes an unwanted outcome but no one is responsible for it, this is a case of **RG***. And if **RG** leads to unjust war, so does **RG***. Hence, if AI shouldn't be used in war because it leads to unjust war, the older technologies also shouldn't be used in war.[89] But it seems implausible that arrows, rifles, catapults, tanks and so on

[88] Notice that my claim isn't that using AI can't lead to violations of the norms of just war. All I claim is that using AI, when it leads to **RG**, doesn't violate the alleged norm of just war. I also don't take position on whether this alleged norm is actually a norm of just war.

[89] One might object that if Soldier takes all the reasonable precautions, Victim's death is an unfortunate accident. Hence, this sort of **RG** can't lead to unjust war. So, those who hold that **RG** leads to unjust war must have something else in mind—i.e., cases where Soldier doesn't take reasonable precautions (or cases where—as I called it above—a 'foul play' is involved). However, if actual consequences of an action *are* relevant for responsibility, in such cases Soldier *is* responsible for Victim's death. Hence, **RG** doesn't occur to begin with. But, if, as I argued above, actual consequences are *not* relevant for responsibility, Soldier is *not* responsible for Victim's death. Hence, **RG** occurs. And if actual consequences are not relevant for responsibility, **RG*** is inevitable just like **RG**. This is

shouldn't be used in war. Indeed, if there's no just war, this isn't implausible. But those who hold that **RG** leads to unjust war neither assume that there's no just war nor take themselves to be arguing against just war theory. On the contrary, it's the norms of just war that they want to uphold (Sparrow, 2007:67; Roff, 2013:352-3).

Two, notice that above we focused exclusively on unwanted consequences and blameworthiness. It should be interesting to note, as it follows from my argument that there's **RG**, that one can also be praiseworthy—indeed *greatly* praiseworthy—for developing or using AI even if the desired outcomes of developing or using AI don't occur. For instance, suppose Soldier deploys AI to save some non-combatants. Even if the AI unexpectedly malfunctions and the mission fails, Soldier can be just as morally praiseworthy as someone who saves the lives of those non-combatants.[90]

4. Conclusion

I argued that AI doesn't generate a new kind of concern about responsibility that the older technologies don't. I then argued that no one's responsibility is affected by what AI might

---

because then neither the consequences of using AI nor the consequences of using older technologies affects anyone's responsibility. And if **RG** leads to unjust war, so does **RG***. Hence, those who hold that **RG** leads to unjust war are again committed to holding that using older technologies also lead to unjust war.

[90] More precisely, when AI malfunctions and the mission fails, Soldier is as praiseworthy as she would have been had the mission not failed. But it doesn't follow that Soldier is as praiseworthy as anyone who saves those non-combatants under any circumstances. Someone (let's call her "Bolder") who actually goes out to the field and saves the non-combatants may be more praiseworthy than Soldier. This is perfectly consistent with my view since we can account for Bolder's being more praiseworthy without appealing to the consequences of her actions—i.e., her success in saving the non-combatants. Plausibly being in the field takes more courage and requires risking more than saving the non-combatants via use of AI which can perfectly well account for why Bolder is more praiseworthy. Hence, had Bolder failed in the end, her degree of praiseworthiness wouldn't have changed. Moreover, it also doesn't follow that given the opportunity to do what Soldier does or what Bolder does, one should prefer the latter since Bolder might be more praiseworthy. This is because once both options are available, there isn't much to be praised for in unnecessarily risking one's life. The overall point here is that we should be careful about the details of the cases we compare.

cause. That is, responsibility gap is inevitable and ubiquitous. However, this isn't worrisome in a way that morally counts against developing or using AI because responsibility gap doesn't lead to any of the alleged problems. On the contrary, I also argued, one would have to hold some fairly implausible views to think otherwise.

Concluding Remarks

I set out to argue that responsibility internalism is true. Responsibility internalism is the idea that basic desert moral responsibility depends only on factors internal to agents. Put differently, the metaphysical ground for basic desert responsibility consists solely of that which is internal to agents.

A big portion of my argument was dedicated to arguably the most pressing concern for responsibility internalism—i.e., the relationship between causal responsibility and moral responsibility. I argued that there is no metaphysically significant relationship between the two. This is because causal responsibility doesn't figure in what best explains neither the degree of moral responsibility nor what makes one morally responsible (or what one is morally responsible for). I also argued that neither the epistemic condition nor the control condition for moral responsibility requires anything external to agents.

As I mentioned in the beginning, I think factors like one's motivation, intention, or care in performing an action are also relevant for moral responsibility. But these are more straightforwardly internal elements, and hence for the most part I left them aside. It would seem then that moral responsibility does not depend on anything external to agents. Hence, I conclude, responsibility internalism is true.

In the final chapter, I employed responsibility internalism to defend a novel position regarding responsibility in the context of AI. I argued that what AI causes affects neither anyone's degree of responsibility nor makes anyone responsible. That is, responsibility gap is

inevitable and ubiquitous. However, I argued, this isn't worrisome in a way that counts against developing or using AI.

References

Alexander, Larry. (2011) Michael Moore and the Mysteries of Causation in the Law. *Rutgers Law Journa*l, 42:301–14.

Alexander, Larry. (2021) Proportionality's Function. *Criminal Law and Philosophy*, 15(3):361-72.

Alexander, Larry, and Ferzan, Kimberly K. (2012) "More of Less" Causation and Responsibility. *Criminal Law and Philosophy*, 6:81-92.

Anderson, Mark B. (2019) Moral Luck as Moral Lack of Control. *The Southern Journal of Philosophy*, 57(1):5-29.

Arkin, Ronald C. (2010) The Case for Ethical Autonomy in Unmanned Systems. *Journal of Military Ethics*, 9(4):332-341.

Balaguer, Mark. (2004) A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will. *Noûs*, 38(3):379–406.

Barker, Kit, and Steele, Jenny. (2015) Drifting Towards Proportionate Liability: Ethics and Pragmatics. *Cambridge Law Journal*, 74, 1:49-77.

Baron, Marcia. (2017) Justification, Excuse, and the Exculpatory Power of Ignorance. *Perspectives on Ignorance from Moral and Social Philosophy*, Peels, R. (ed.), (pp.53-76), New York: Routledge.

Bebe, Helen. (2013) Legal Responsibility and Scalar Causation. *Jurisprudence*, 4(1):102-8.

Bernstein, Sara. (2016) Causal and Moral Indeterminacy. *Ratio,* 29(4): 434-447.

Bernstein, Sara. (2017) Causal Proportions and Moral Responsibility. *Oxford Studies in Agency and Responsibility*, David Shoemaker (ed.), (pp.165-182), Oxford: Oxford University Press.

Bernstein, Sara. (2019) Moral Luck and Deviant Causation. *Midwest Studies in Philosophy*, 43 (1):151-161.

Braham, Matthew, and van Hees, Martin. (2009) Degree of Causation. *Erkenntnis,* 71(3): 323-344.

Bunzl, Martin. (1979) Causal Overdetermination. *The Journal of Philosophy*, 76:134-150.

Burri, Susanne. (2017) What Is the Moral Problem with Killer Robots? *Who Should Die? The Ethics of Killing in War*, (pp.163-85), Jenkins, R., Robillard, M., Strawser, B. J. (eds.), New York: OUP.

Chisholm, Roderick M. (1966) Freedom and Action. *Freedom and Determinism*, Keith Lehrer (ed.), (pp.11-44), New York: Random House.

Chockler, Hana, and Halpern, Joseph Y. (2004) Responsibility and Blame: A Structural-Model Approach. *Journal of Artificial Intelligence Research*, 22:93-115.

Clarke, Randolph. (1993) Toward a Credible Agent-Causal Account of Free Will. *Noûs*, 27: 191–203.

Clarke, Randolph. (2017) Ignorance, Revision, and Commonsense. *Responsibility: The Epistemic Condition*, Robichaud, P., and Wieland, J. W. (eds.), (pp.233-51), Oxford: OUP.

Cyr, Taylor W. (2019) Why Compatibilists Must Be Internalists. *The Journal of Ethics*, 23:473-84.

Cyr, Taylor W. (forthcoming) Taking Hobart Seriously. *Philosophia*.

Danaher, John. (2016) Robots, Law and The Retribution Gap. *Ethics and Information Technology*, 18:299–309.

Demetriou, Kristin. (2010) The Soft-Line Solution to Pereboom's Four-Case Argument. *Australasian Journal of Philosophy*, 88(4):595-617.

Demirtas, Huzeyfe. (2022a) Against Resultant Moral Luck. *Ratio*, 35(3):225-235.

Demirtas, Huzeyfe. (2022b) Causation Comes in Degrees. *Synthese*, 200, 1–17.

Demirtas, Huzeyfe (2022c) Moral Responsibility Is Not Proportionate to Causal Responsibility. *Southern Journal of Philosophy*, 60(4):570-91.

Dowe, Phil. (2000) *Physical Causation*. Cambridge: Cambridge University Press.

Dowe, Phil. (2013) Moore's Account of Causation and Responsibility, and the Problem of Omissive Overdetermination. *Jurisprudence*, 4(1):115-20.

Eells, Ellery. (1991) *Probabilistic Causality*. Cambridge: Cambridge UP.

Ekstrom, Laura W. (2019) Toward a Plausible Event-Causal Indeterminist Account of Free Will. *Synthese*, 196:127–144.

Feinberg, Joel. (1968) Collective Responsibility. *The Journal of Philosophy*, 65:674-688.

Feldman, Richard. (2003) *Epistemology*. Upper Saddle River, NJ: Prentice Hall.

Fischer, John. M. (1994) *The Metaphysics of Free Will: An Essay on Control*, Oxford: Blackwell Publishers.

Fischer, John M. (2007) Compatibilism. *Four Views on Free Will*, Fischer et al. (eds.), (pp.44-84), Hoboken, NJ: Wiley-Blackwell.

Fischer, John. M., and Ravizza, Mark. (1998) *Responsibility and Control: A theory of Moral Responsibility*. New York: Cambridge University Press.

Fitelson, Brandon, Hitchock, Christopher. (2011) Probabilistic Measures of Causal Strength. *Causality in the Sciences*, Illari, P. M., Russo, F., Williamson J. (eds.), (pp.600-27), New York, OUP.

Fitzpatrick, William J. (2017) Unwitting Wrongdoing, Reasonable Expectations, and Blameworthiness. *Responsibility: The Epistemic Condition*, Robichaud, P., and Wieland, J. W. (eds.), (pp.29-46), Oxford: OUP.

Frankfurt, Harry. (1971) Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1):5–20.

Frankfurt, Harry. (2002) Reply to John Martin Fischer. *Contours of Agency: Essays on Themes from Harry Frankfurt*, S. Buss, and L. Overton (eds.), (pp.27-31), Cambridge, MA: The MIT Press.

Goetze, Trystan. (2022) Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*.

Gerstenberg, Tobias, Ullman, Tomer D., Nagel, Jonas, Kleiman-Weiner, Max, Lagnado, David A., and Tenenbaum, Joshua B. (2018) Lucky or clever? From Expectations to Responsibility Judgments. *Cognition*, 177:122-41.

Gerstenberg, Tobias, Goodman, Noah D., Lagnado, David A., and Tenenbaum, Joshua B. (forthcoming) A Counterfactual Simulation Model of Causal Judgments for Physical Events. *Psychological Review*.

Ginet, Carl. (1997) Freedom, Responsibility, and Agency. *Journal of Ethics*, 1:85–98.

Goldman, Alvin, and Bob Beddor. (2021) Reliabilist Epistemology. *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/reliabilism/>.

Griffith, Meghan. (2010) Why Agent-Caused Actions are Not Lucky. *American Philosophical Quarterly*, 47:43–56.

Haji, Ishtiyaque. (2008) *Incompatibilism's Allure: Principal Arguments for Incompatibilism*. Ontario: Broadview Press.

Hales, Steven. (2015) A Problem for Moral Luck. *Philosophical Studies*, 172(9):2385–403.

Hall, Ned. (2000) Causation and the Price of Transitivity. *Journal of Philosophy*, 97:198–222.

Hall, Ned. (2004) Two Concepts of Causation. *Causation and Counterfactuals*, Collins, J., Hall, N., Paul, L. A. (eds.), (pp.225-276), MIT Press, Cambridge, MA.

Hartman, Robert J. (2016) Against Luck-Free Moral Responsibility. *Philosophical Studies*, 173:2845-65.

Hartman, Robert J. (2017) *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness*. New York: Routledge.

Hartman, Robert J. (Forthcoming) Free Will and Moral Luck. *A Companion to Free Will*, Campbell, J., Mickelson, K. M., and White, V. A. (eds). Wiley-Blackwell.

Hieronymi, Pamela. (2014) Reflection and Responsibility. *Philosophy and Public Affairs*, 42(1):3-41.

Himmelreich, Johannes. (2019) Responsibility for Killer Robots. *Ethical Theory and Moral Practice*, 22:731–747.

Himmelreich, Johannes, and Köhler, Sebastian. (2022) Responsible AI Through Conceptual Engineering. *Philosophy & Technology*, 35(3):1-30.

Hitchcock, Christopher. (2018) Probabilistic Causation. *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/causation-probabilistic/>.

Hobart, R. E. (1934) Free Will as Involving Determination and Inconceivable Without It. *Mind*, 43:1-27.

Hunt, David. (2000) Moral Responsibility and Unavoidable Action. *Philosophical Studies*, 97:195–227.

Hevelke, Alexander, and Nida-Rümelin, Julian. (2015) Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. *Science and Engineering Ethics*, 21(3):619-630.

Icard, Thomas F., Kominsky, Jonathan F., and Knobe, Joshua. (2017). Normality and Actual Causal Strength. *Cognition*, 161:80-93.

Kaiserman, Alex. (2016) Causal Contribution. *Proceedings of the Aristotelean Society*, 116:387-394.

Kaiserman, Alex. (2017) Partial Liability. *Legal Theory*, 23:1-26.

Kaiserman, Alex. (2018) 'More of a Cause': Recent Work on Degrees of Causation and Responsibility. *Philosophy Compass*, 13(7):e12498.

Kaiserman, Alex. (2021) Responsibility and the 'Pie Fallacy'. *Philosophical Studies*, 178:3597–3616.

Kane, Robert. (2016) On the Role of Indeterminism in Libertarian Free Will. *Philosophical Explorations*, 19:2-16.

Kershnar, Stephen. (2018) *Total Collapse: The Case Against Responsibility and Morality*. Cham, Switzerland: Springer.

Khoury, Andrew. (2013) Synchronic and Diachronic Responsibility. *Philosophical Studies*, 165:735-52.

Khoury, Andrew. (2014) Manipulation and Mitigation. *Philosophical Studies*, 168:283–294.

Khoury, Andrew. (2017) Individual and Collective Responsibility. *Reflections on Ethics and Responsibility: Essays in Honor of Peter A. French*, (pp.1-20), Goldberg, Z. J. (ed.). Springer International Publishing.

Khoury, Andrew. (2018) The Objects of Moral Responsibility. *Philosophical Studies*, 175:1357-81.

Khoury, Andrew C. and Matheson, Benjamin. (2018) Is Blameworthiness Forever? *Journal of the American Philosophical Association*, 4(2):204-24.

Kiener, Maximilian. (2022) Can We Bridge AI's Responsibility Gap at Will? *Ethical Theory and Moral Practice*, 25(4):575-593.

Kneer, Markus, and Machery, Edouard. (2019) No Luck for Moral Luck. *Cognition*, 182:331-348.

Korb, Kevin B., Nyberg, Erik, and Hope, Lucas. (2011) A New Causal Power Theory. *Causality in the Sciences*, Illari, P. M., Russo, F., Williamson J. (eds.), (pp.628-52), New York, OUP.

Köhler, Sebastian, Roughley, Neil, and Sauer, Hanno. (2017). Technologically Blurred Accountability? Technology, Responsibility Gaps and The Robustness of Our Everyday Conceptual Scheme. *Moral Agency and the Politics of Responsibility*, (pp. 51–67), C. Ulbert, P. Finkenbusch, E. Sondermann, & T. Debiel (eds.). Routledge.

Königs, Peter. (forthcoming) Artificial Intelligence and Responsibility Gaps: What is The Problem? *Ethics and Information Technology*.

Kvart, Igal. (2002) Probabilistic Cause and the Thirsty Traveler. *Journal of Philosophical Logic*, 31:139–79.

Langenhoff, Antonia F., Wiegmann, Alex, Halpern, Joseph Y., Tenenbaum, Joshua B., and Gerstenberg, Tobias. (forthcoming) Predicting Responsibility Judgments from Dispositional Inferences and Causal Attributions. *Cognitive Psychology*.

Lagnado, David, A., Gerstenberg, Tobias, and Zultan, Ro'i. (2014) Causal Responsibility and Counterfactuals. *Cognitive Science*, 37:1036-1073.

Levy, Neil. (2005) The Good, the Bad, and the Blameworthy. *Journal of Ethics and Social Philosophy*, 1(2):2-16.

Levy, Neil. (2011) *Hard Luck: How Luck Undermines Free Will and Responsibility*. Oxford: Oxford University Press.

Levy, Neil. (2017) Methodological Conservatism and the Epistemic Condition. *Responsibility: The Epistemic Condition*, Robichaud, P., and Wieland, J. W. (eds.), (pp.252-265), Oxford: OUP.

Lewis, David. (2004) Causation as Influence. *Causation and Counterfactuals*, Collins, J., Hall, N., Paul, L. A. (eds.), (pp.75-106), MIT Press, Cambridge, MA.

Matthias, Andreas. (2004) The Responsibility Gap: Ascribing Responsibility for The Actions of Learning Automata. *Ethics and Information Technology*, 6:175–183.

McKenna, Michael. (2008). A Hard-line Reply to Pereboom's Four-Case Manipulation Argument. *Philosophy and Phenomenological Research*, 77: 142-159.

McLaughlin, James A. (1925) Proximate cause. *Harvard Law Review*, 39:149–99.

Mele, Alfred. (2009) Moral Responsibility and History Revisited. *Ethical Theory and Moral Practice*, 12:463–475.

Mele, Alfred. (2016) Moral Responsibility: Radical Reversals and Original Designs. *Journal of Ethics*, 20:69–82.

Mele, Alfred. (2019) *Manipulated Agents: A Window to Moral Responsibility*. New York: Oxford University Press.

Miller, David. (2001) Distributing Responsibilities. *The Journal of Political Philosophy,* 9(4):453-71.

Mitchell-Yellin, Benjamin. (2015) The Platonic Model: Statement, Clarification and Defense. *Philosophical Explorations*, 18: 378–92.

Moore, Michael S. (2009) *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*, Oxford: Oxford University Press.

Moore, Michael S. (2011) Causation Revisited. *Rutgers Law Journal*, 42:451-509.

Moore, Michael S. (2012) Moore's Truths About Causation and Responsibility: A Reply to Alexander and Ferzan. *Criminal Law and Philosophy*, 6:445-62.

Moore, Michael S. (2013) Author's Reply, *Jurisprudence*, 4(1):121-37.

Mumford, Stephen. (2013) Causes for Laws. *Jurisprudence*, 4(1):109-14.

Müller, Vincent C. (2016) Autonomous Killer Robots Are Probably Good News. *Drones and Responsibility: Legal, Philosophical and Socio-technical Perspectives on The Use of Remotely Controlled Weapons*, (pp.67-81), Di Nucci, E., and Santonio de Sio, F. (eds.). London: Ashgate.

Müller, Vincent C., (2021) Ethics of Artificial Intelligence and Robotics. *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/>.

Nagel, Thomas. (1979) *Mortal Questions*, New York: Cambridge University Press.

Nelkin, Dana K. (2019) Moral Luck. *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/moral-luck/>.

Northcott, Robert. (2005a) Comparing Apples with Oranges. *Analysis*, 65(1):12–8.

Northcott, Robert. (2005b) Pearson's Wrong Turning: Against Statistical Measures of Causal Efficacy. *Philosophy of Science*, 72(5):900–912.

Northcott, Robert. (2008) Weighted Explanations in History. *Philosophy of the Social Sciences*, 38 (1):76-96.

Northcott, Robert. (2013) Degree of Explanation. *Synthese*, 190(15):3087–105.

Nyholm, Sven. (2018a) Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. *Science and Engineering Ethics*, 24(4):1201-1219.

Nyholm, Sven. (2018b) The Ethics of Crashes with Self-Driving Cars: A Roadmap, II. *Philosophy Compass*, 13(7):e12506.

O'Connor, Timothy and Christopher Franklin. (2021) Free Will. *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/freewill/>.

Palmer, David. (2016) Goetz on the Noncausal Libertarian View of Free Will. *Thought*, 5:99-107.

Paul, Laurie A., Hall, Edward J. (2013) *Causation: A User's Guide*. Oxford: Oxford University Press.

Pappas, George. (2017) Internalist vs. Externalist Conceptions of Epistemic Justification. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/justep-intext/>.

Pearson, Richard N. (1980) Apportionment of Losses Under Comparative Fault Laws—An Analysis of the Alternatives. *Louisiana Law Review*, 40(2):323-72.

Peels, Rick. (2015) A Modal Solution to the Problem of Moral Luck. *American Philosophical Quarterly*, 52(1):73–87.

Pereboom, Derk. (1995) Determinism Al Dente. *Noûs*, 29(1):21–45.

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

Petersson, Björn. (2013) Co-responsibility and Causal Involvement. *Philosophia*, 41(3):847-866.

Purves, Duncan, and Jenkins, Ryan. (2016) Robots and Respect: A Response to Robert Sparrow. *Ethics and International Affairs*, 30(3):391-400.

Roff, Heather M. (2013) Responsibility, Liability, and Lethal Autonomous Robots. *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century*, (pp.352-64), Allhoff, F., Evans, N., and Henschke, A. (eds.). Routledge.

Rosen, Gideon. (2004) Skepticism about Moral Responsibility. *Philosophical Perspectives*, 18: 295–313.

Rosen, Gideon. (2008) Kleinbart the Oblivious and Other Tales of Ignorance and Responsibility. *Journal of Philosophy*, 105(10):591-610.

Rosen, Gideon. (2011) Causation, Counterfactual Dependence and Culpability: Moral Philosophy in Michael Moore's Causation and Responsibility. *Rutgers Law Journal*, 42:405-434.

Rudy-Hiller, Fernando. (2017) A Capacitarian Account of Culpable Ignorance. *Pacific Philosophical Quarterly*, 98(S1): 398–426.

Rudy-Hiller, Fernando. (2018) The Epistemic Condition for Moral Responsibility. *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/moral-responsibility-epistemic/>.

Rudy-Hiller, Fernando. (2021) It's (Almost) All About Desert: On the Source of Disagreements in Responsibility Studies. *The Southern Journal of Philosophy*, 59(3):386-404.

Sars, Nicholas. (2022) Engineering Responsibility. *Ethics and Information Technology*, 24(3):1-10.

Sartorio, Carolina. (2004) How to be Responsible for Something Without Causing It. *Philosophical Perspectives*, 18(1):315–336.

Sartorio, Carolina. (2007) Causation and Responsibility. *Philosophy Compass*, 2(5):749-65.

Sartorio, Carolina. (2010) Causation and Responsibility by Michael S. Moore. *Mind*, 119:830-8.

Sartorio, Carolina. (2012) Resultant Luck. *Philosophy and Phenomenological Research*, 84(1):63-86.

Sartorio, Carolina. (2015a) A New Form of Moral Luck? *Agency, Freedom, And Moral Responsibility*, Buckareff, A., Moya, C., Rosell, S. (eds.), (pp.134-149), New York: Palgrave-Macmillan.

Sartorio, Carolina. (2015b) Resultant Luck and the Thirsty Traveler. *Methode*, 4:153–71.

Sartorio, Carolina. (2016) Causation and Free Will. New York: Oxford University Press.

Sartorio, Carolina. (2020) More of a Cause? *Journal of Applied Philosophy*, 37(3):346-63.

Schaffer, Jonathan. (2003) Overdetermining Causes. *Philosophical Studies*, 114:23-45.

Schaffer, Jonathan. (2016) The Metaphysics of Causation. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/>.

Shabo, Seth. (2010) Uncompromising Source Incompatibilism. *Philosophy and Phenomenological Research*, 80:349–83.

Sher, George. (2009) *Who Knew? Responsibility without Awareness* , New York: Oxford University Press.

Shoemaker, David. (2003) Caring, Identification, and Agency. *Ethics*, 114:88–118.

Shoemaker, David. (2011) Attributability, Answerability, and Accountability: Toward A Wider Theory of Moral Responsibility. *Ethics*, 121:602-32.

Shoemaker, David. (2012) Responsibility without Identity. *Harvard Review of Philosophy*, 18(1):109-132.

Shoemaker, David. (2015) *Responsibility from the Margins*. Oxford, United Kingdom: Oxford University Press.

Shoemaker, David (2020) Responsibility: The State of the Question. Fault Lines in the Foundations. *The Southern Journal of Philosophy*, 58(2):205–37.

Simpson, Thomas W., & Müller, Vincent C. (2016) Just War and Robots' Killings. *Philosophical Quarterly*, 66(263):302-22.

Smiley, Marion. (2022) Collective Responsibility. *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2022/entries/collective-responsibility/>.

Smith, Angela. (2012) Attributability, Answerability, and Accountability: In Defense of a Unified Account. *Ethics*, 122:575-89.

Smith, Angela. (2015) Responsibility as Answerability. *Inquiry*, 58(2):99-126.

Sparrow, Robert. (2007) Killer Robots. *Journal of Applied Philosophy*, 24:62–77.

Sparrow, Robert. (2016) Robots and Respect: Assessing the Case Against Autonomous Weapon Systems. *Ethics and International Affairs*, 30(1):93-116.

Sprenger, Jan. (2018) Foundations of Probabilistic Theory of Causal Strength. *Philosophical Review*, 127(3):371-98.

Stump, Eleonore. (1999) Alternative Possibilities and Moral Responsibility: The Flicker of Freedom. *The Journal of Ethics*, 3(3):299-324.

Sverdlik, Steven. (1988) Crime and Moral Luck. *American Philosophical Quarterly*, 25(1):79-86.

Tadros, Victor. (2018) Causal Contributions and Liability. *Ethics*, 128:402-31.

Talbert, Matthew. (2019) Moral Responsibility. *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>.

Thomson, Judith J. (1993) Morality and Bad Luck. *Moral Luck*, (pp.195-216), Statman, D. (ed.), Albany: State University of New York Press.

Tiefensee, Christine. (2019) Why Making No Difference Makes No Moral Difference, *Demokratie und Entscheidung*, (pp.231-244) Marker K., Schmitt A., Sirsch J. (eds.), Springer VS, Wiesbaden.

Tierney, Hannah. (2013) A Maneuver Around The Modified Manipulation Argument. *Philosophical Studies*, 165 (3):753-763.

Tigard, Daniel W. (2021) There Is No Techno-Responsibility Gap. *Philosophy & Technology*, 34:589–607.

Vallentyne, Peter. (2008) Brute Luck and Responsibility. *Politic, Philosophy and Economics*, 7(1):57-80.

van Inwagen, Peter. (1983) *An Essay on Free Will*. Oxford: Clarendon.

van Inwagen, Peter (2015) *Metaphysics*. Boulder, CO: Westview Press.

Vargas, Manuel. (2012) Why the Luck Problem Isn't. *Philosophical Issues*, 22, 419–436.

Vincent, Nicole A. (2011) A Structured Taxonomy of Responsibility Concepts. *Moral Responsibility: Beyond Free Will and Determinism*, (pp.15-35), Vincent, N., Poel, I. and Hoven, J. (eds.), Dordrecht New York: Springer.

Wakabayashi, Daisuke. (2018) Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. *The New York Times*, March 19, 2018. Accessed December 20, 2022. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html.

Watson, Gary. (1975) Free Agency. *Journal of Philosophy*, 72(April):205-20.

Watson, Gary. (1999) Soft Libertarianism and Hard Incompatibilism. *Journal of Ethics*, 3: 353–368.

Whittington, Lee J. (2014) Getting Moral Luck Right. *Metaphilosophy*, 45(4–5): 654–67.

Widerker, David. (2018) In Defense of Non-Causal Libertarianism. *American Philosophical Quarterly*, 55(1):1-14.

Wieland, Jan Willem. (2017) Introduction: The Epistemic Condition. *Responsibility: The Epistemic Condition*, (pp.1-28), Robichaud, P., and Wieland, J. W. (eds.), Oxford: OUP.

Williams, Bernard. (1981) Moral Luck. *Moral Luck* (pp.20-39), Williams, B. (ed.), Cambridge: Cambridge University Press.

Wolf, Susan. (1987) Sanity and the Metaphysics of Responsibility. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, (pp.46-62), Schoeman, F. (ed.), Cambridge: Cambridge University Press.

Wolf, Susan. (1990) *Freedom within Reason*. Oxford: Oxford University Press.

Wolf, Susan. (2001) The Moral of Moral Luck. *Philosophic Exchange*, 31: 4–19.

Zagzebski, Linda. (2000) Does Libertarian Freedom Require Alternate Possibilities? *Philosophical Perspectives*, 14:231–48.

Zimmerman, Michael J. (1985) Sharing Responsibility, *American Philosophical Quarterly* 22:115-22.

Zimmerman, Michael J. (1987) Luck and Moral Responsibility. *Ethics*, 97(2):374-386.

Zimmerman, Michael. (1997) Moral Responsibility and Ignorance. *Ethics*, 107(3): 410–26.

Zimmerman, Michael J. (2006) Moral Luck: A Partial Map. *Canadian Journal of Philosophy*, 36(4):585-608.

Zimmerman, Michael J. (2015) Varieties of Moral Responsibility, *The Nature of Moral Responsibility: New Essays*, (pp.45-64), Clarke, R., McKenna, M., Smith, A. M. (eds.), New York: Oxford University Press.

Curriculum Vitae

Huzeyfe Demirtas

CONTACT

www.huzeyfedemirtas.com

Department of Philosophy
Syracuse University
541 Hall of Languages, Syracuse,
NY 13210

SPECIALIZATION

Ethics, Applied Ethics (esp. Environmental Ethics)

COMPETENCIES

Epistemology, Metaphysics, Classical Islamic Philosophy, Political Philosophy

EDUCATION

Syracuse University 2016- 2023
PhD in Philosophy
Dissertation: *Responsibility Internalism & Responsibility for AI*
Committee: Sara Bernstein, Ben Bradley (primary), Mark Heller, Hille Paakkunainen

SUNY Fredonia 2015-2016
Postbaccalaureate in Philosophy

Firat University (*Turkey*) 2004-2009
BS in Computer Science Teaching

PUBLICATIONS

'Moral Responsibility Is Not Proportionate to Causal Responsibility'
The Southern Journal of Philosophy 2022

'Against Resultant Moral Luck' 2022
Ratio

'Causation Comes in Degrees' 2022
Synthese

PUBLIC PHILOSOPHY

'Epistemic Injustice' 2020
1000-Word Philosophy: An Introductory Anthology

TALKS

Speaker *(*=refereed)*
*'Drawing A Line, Rejecting Resultant Moral Luck Alone' Apr 2023
American Philosophical Association, Pacific Division Meeting 2023

*'Drawing A Line, Rejecting Resultant Moral Luck Alone' Feb 2023
Free Will, Moral Responsibility, and Agency
Florida State University

*'Take a Stand, You Don't Have to Make a Difference' Jan 2023
2022-23 Young Philosophers Read-Ahead Conference
DePauw University

*'Take a Stand, You Don't Have to Make a Difference' Jan 2023
International Society for Environmental Ethics,
American Philosophical Association, Eastern Division Meeting 2023

'Drawing A Line, Rejecting Resultant Moral Luck Alone' Oct 2022
ABD Workshop Series, Syracuse University

*'Take a Stand, You Don't Have to Make a Difference' Sep 2022
You Philosophers Lecture Series, DePauw University

*'Wrong but Praiseworthy, Right but Blameworthy' June 2022
Rightness, Ignorance, Uncertainty, and Praise Workshop
University of Southern California

*'Wrong but Praiseworthy, Right but Blameworthy' Apr 2022
72nd Annual Meeting of the New Mexico Texas Philosophical Society
Baylor University

'Wrong but Praiseworthy, Right but Blameworthy' Feb 2022
ABD Workshop Series, Syracuse University

*'Causation Comes in Degrees' Jan 2022
American Philosophical Association, Eastern Division Meeting

*'Causation Comes in Degrees' Sep 2021
Society for the Metaphysics of Science, 6th Annual Conference

*'Against Resultant Moral Luck' July 2021
Summer School on Causation and Responsibility,
University of Bern

*'Against Resultant Moral Luck' Apr 2021
Great Lakes Philosophy Conference—Ethics in Action,
Siena Heights University

*'Moral Responsibility Is Not Proportionate to Causal Responsibility'
American Philosophical Association, Eastern Division Meeting Jan 2021

'Against Resultant Moral Luck' Nov 2020
Philosophical Society of Fredonia, SUNY Fredonia

*'Against Resultant Moral Luck' July 2020
94th Joint Session of the Aristotelian Society and the Mind Association,
University of Kent

*'Causal Contributions and Moral Responsibility' Mar 2020
Midsouth Philosophy Conference [Canceled],
Rhodes College

'Moral Responsibility Is Not Proportionate to Causal Responsibility'
ABD Workshop Series 2020 Feb 2020
Syracuse University

*'Causal Contributions and Moral Responsibility' Nov 2019
AGENT, Ethics and Normativity Talks,
University of Texas at Austin

*'Against Proportionality Luck' June 2019
International Conference on Ethics, University of Porto

*'Against Proportionality Luck' Mar 2019
20th Annual Pitt-CMU Graduate Student Philosophy Conference,
University of Pittsburgh & Carnegie Mellon University

*'Stocker's Schizophrenia, Alienation, and a Solution' Apr 2018
Fundamentality in Philosophy, The 7th International Philosophy Graduate Conference,
Central European University, Budapest

*'Against Reliabilism: In the Face of Skepticism' May 2017
Northwest Student Philosophy Conference,
Western Washington University

Commentator
On Itamar Weinshtock Saadon's 'Responsibility, Causation, and Reversing the Order of Explanation'
Syracuse Graduate Philosophy Conference Mar 2023

On Joshua Tignor's 'Theorizing About Moral Responsibility As Such'
ABD Workshop Series 2021, Feb 2023
Syracuse University

On Jules Salomone-Sehr's 'Complicity: A Minimalist Account for Our Maximally Messy Social World' July 2022
Vancouver Summer Philosophy Conference

On Hannah Winckler-Olick's 'Simone de Beauvoir on Value-Creation as a Mode of Complicity'
Centennial Conference of the Creighton Club Apr 2022

On Peter Zuk's 'Reconciling Experiential Theories of Pleasure'

72nd Annual Meeting of the New Mexico Texas Philosophical Society,
Baylor University Apr 2022

On David Sackris and Rasmus Rosenberg Larsen's 'Are There Moral Judgements?' Jan 2022
American Philosophical Association, Eastern Division Meeting

On Joshua Tignor's 'Moral Growth and Moral Responsibility'
ABD Workshop Series 2021, Oct 2021
Syracuse University

On Alex Kaiserman's 'Responsibility and the 'Pie Fallacy'' July 2021
Summer School on Causation and Responsibility,
University of Bern

On Perry Hendricks's 'The Impairment Argument Reconsidered'
Syracuse Graduate Philosophy Conference Apr 2021

On Caner Turan's 'On Greene's Evolutionary Challenge to Deontological Ethics'
Syracuse Graduate Philosophy Conference Mar 2019

TEACHING

Syracuse University

—Lead Instructor:

PHI191: The Meaning of Life	Fall 2022
PHI394: Environmental Ethics (X2)	Spring 2022/23
PHI251: Logic (X3)	Spring 2020, Summer 2021/22
PHI383: Free Will	Spring 2021
PHI200: Happiness and Meaning in Life	Winter 2021
PHI197: Human Nature	Fall 2020
PHI107: Theories of Knowledge and Reality	Summer 2020
PHI192: Introduction to Moral Theory	Fall 2019

—Teaching Assistant:

Human Nature (Christopher Noble)	Fall 2021
Theories of Knowledge and Reality (Janice Dowell)	Spring 2019
Logic (Mark Heller)	Fall 2018
Introduction to Moral Theory (David Sobel)	Spring 2018
Introduction to Moral Theory (Hille Paakkunainen)	Fall 2017
Human Nature (Neelam Sethi)	Spring 2017
Theories of Knowledge and Reality (Robert Van Gulick)	Fall 2016

AWARDS

Syracuse University 2022
Summer Research Fellowship

Syracuse University 2021
Outstanding Teaching Assistant Award

SUNY Fredonia
The Philosophical Society, Student Achievement Award 2016

SERVICE

Referee for:
American Philosophical Quarterly, Australasian Journal of Philosophy, Ergo, European Journal of Philosophy, Synthese

Senator
Syracuse Graduate Student Organization 2020 – 2021

Co-Organizer
Syracuse Graduate Philosophy Conference 2020

GRADUATE COURSEWORK

Ethics *(*=audit)*
Moral and Political Philosophy (Hille Paakkunainen)
Constructivism in Metaethics (Hille Paakkunainen)
Anti-Realism and Pragmatism in Ethics (Nate Sharadin)
Anti-Theory in Ethics (Independent study with Hille Paakkunainen)
Ethics of Nudging (Independent study with Ben Bradley)
*Motivation (Hille Paakkunainen)
*Animal Ethics (Ben Bradley)
*Free Will (Mark Heller)
*Prudence (Ben Bradley)

Epistemology *(*=audit)*
Topics in Contemporary Epistemology (Nate Sharadin)
Language, Epistemology, Mind, Metaphysics (K.McDaniel, K.Edwards)
*Epistemology (Hille Paakkunainen)

Metaphysics
Beyond the Modal: Essence and Potentiality (Kris McDaniel)
Metaphysics of Ethics (Ben Bradley, Kris McDaniel)

Political Philosophy
Justice and Equality (Ken Baynes)
Philosophy of Social Sciences (Ken Baynes)

History of Philosophy
History of Philosophy (Frederick C. Beiser)
Classical Arabic Philosophy (Kara Richardson)

Logic and Language
Logic and Language (Michael Rieppel)
Concepts (Kevan Edwards)

LANGUAGES

English, Turkish (Native), Arabic (Reading, Intermediate)