Why Unavoidable Moral Wrongdoing Is Impossible

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Philosophy

at

The University of Waikato

by

Aaron Lane

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2025

Abstract

We can be faced with situations in which we act in accordance with the all-things-considered morally right course of action, and yet nevertheless experience an emotion of self-directed distress at the fact that we have so acted. Moreover, we have the intuition that this emotion is rational or appropriate. Literature is scattered with such examples, and they crop up not infrequently in life.

The fact that agents experience justified distress in such situations has been taken as evidence that unavoidable moral wrongdoing is possible. However, if this is the case, then the compelling intuition that an all-things-considered morally right action cannot be wrong must be false.

In this thesis, I offer a way of reconciling these two compelling intuitions that denies the possibility of unavoidable moral wrongdoing. I argue that existing attempts to dissolve this contradiction by arguing in favour of the possibility of unavoidable moral wrongdoing are unsatisfactory; and I contend that, by utilising Susan Wolf's ideas on the "nameless virtue" (2001), developed in order to solve the problem of resultant moral luck, we can arrive at a more compelling method of dissolving the apparent contradiction. This solution is one that allows us to retain the intuition that a right act cannot also be wrong, while simultaneously allowing for the rationality of emotion of self-directed distress agents experience in situations of apparent unavoidable moral wrongdoing.

Acknowledgements

I would like to thank my chief supervisor, Liezl van Zyl, for her generous support and advice throughout the research and writing process. I'd also like to thank Nick Munn, my second supervisor, for his helpful comments and suggestions. Finally, I'd like to acknowledge the University of Waikato Philosophy department more broadly for creating a collegial, welcoming environment. I've enjoyed being a part of it for these last three years.

Table of Contents

Introduction

Unavoidable moral wrongdoing (UMW) refers to situations in which, even if an agent makes the correct all-things-considered decision, she does wrong. This thesis is an investigation into whether such wrongdoing is possible.

It is sometimes argued that there are genuine (or irresolvable) moral dilemmas.[1] Some philosophers consider a moral dilemma to be genuine when neither of an agent's conflicting obligations is overridden. Sinnot-Armstrong, for instance, argues that a genuine moral dilemma is a situation "where there is a moral requirement for an agent to adopt each of two alternatives, and the agent cannot adopt both, but neither moral requirement is overridden in a morally relevant way." (1985, p. 322)[2] Others consider a genuine moral dilemma to be a situation in which there is no all-things-considered morally correct action for an agent to take.[3] On this view, an agent in a symmetrical dilemma (a dilemma in which each obligation is morally identical in every respect) is not in a genuinely dilemmatic situation unless the disjunctive requirement—the requirement that the agent simply choose to act on one of the two obligations, at random or on the basis of some non-moral reason-–does not apply. This is because, if the disjunctive requirement does apply, then there *is* an all-things-considered morally correct course of action.[4]

In this thesis, I will assume that the latter is correct—that a genuine moral dilemma occurs when an agent has no all-things-considered morally correct course of action open to her. I will further assume that there are no such situations—there is always an all-things-considered morally correct course of action for an agent to pursue.[5] In other words, I assume that if an agent faces a choice between A and B, one of the following is true:

1. The agent has an obligation to do A.
2. The agent has an obligation to do B.
3. The agent has an obligation to do either A or B.

---

[1] See Williams (1973), Marcus (1980), Foot (1983), Sinnot-Armstrong (1985, 1987), and Tessman (2015.)

[2] For a more recent example, see Bagnoli (2012), who joins Sinnot-Armstrong in arguing that where two options are of equal moral value and the agent must decide on some arbitrary basis such as the toss of a coin, because there is no morally salient reason for choosing one option over the other, the dilemma is unresolved.

[3] Daniel Statman (1995), for instance, takes this view.

[4] See, for instance, Brink (1994) and Statman (1995.)

[5] Extensive and convincing arguments for this position have been made by Christopher Gowans (1994, Ch. 3-4) and Daniel Statman (1995). I outline Gowans's arguments briefly in Chapter Two.

1 and 2 obtain when the reasons to do A outweigh the reasons to do B or vice versa. 3 obtains when neither the reasons to do A nor the reasons to do B outweigh the other; in these (rare) situations, there is a disjunctive requirement, with the all-things-considered morally correct course of action being for the agent to do A or B, but not both.[6] Therefore, the broad starting point for this thesis is the following question: assuming that there is an all-things-considered morally right action for an agent in a given situation, is it nevertheless possible for the agent to do wrong?

On one hand, that it could be the case seems dubious.[7] A strong intuition supports this view. It seems conceptually incoherent to hold that one can do moral wrong in doing what one morally ought to do. To say that an action is right seems logically to entail that the action is not wrong. Michael Walzer captures the implausibility of such a possibility well: "[H]ow can it be wrong to do what is right? How can we get our hands dirty doing what we ought?" (1973, p. 164) We can state this intuitive principle as follows:

> **An action that is all-things-considered morally right cannot also be morally wrong (RCW).**[8]

The puzzlement—almost incredulity—that is evident in Walzer's interrogatives likely stems from the fact that, when one considers this intuition, it seems obviously correct, perhaps even an analytic truth. If an act is right, all-things-considered, then it appears clearly true that it cannot also be wrong. To say that it can be is as implausible as claiming that something can be both entirely red and entirely blue at the same time, or that 15 is less than 14.

On the other hand, when we move away from the abstract and instead consider specific instances from moral life, it does appear that there can be situations in which an agent appropriately feels moral distress even if he acts for the best, and this calls into question RCW. As Daniel Statman puts it, "moral experience seems to testify in favor of the existence of moral dilemmas, [and so] the

---

[6] This set of possibilities captures what Vallentyne (1987) terms "obligation dilemmas" (where an agent has prima facie obligations to do both A and B) and "prohibition dilemmas" (where an agent has prima facie obligations not to do A or B.) This is because, in a "prohibition dilemma" situation, either one of the two prohibited actions will seem to represent less of a moral breach than the other, and so the all things considered morally correct action is to perform that action; or, if they are both equally morally problematic, then the action's obligation is to do one or the other.

[7] Philosophers who have held that unavoidable moral wrongdoing is impossible include Sidgwick (1901), Ross (1930), Hare (1987), McConnell (1996), and Nielsen (2000.)

[8] De Wijze states a similar principle as follows (although he does not endorse it): "If A is obligatory, it cannot also be forbidden." (2022, p. 197) He terms it the Principle of Deontic Consistency (PDC.)

onus rests on whoever wishes to deny their existence." (1995, p. 29)[9] In these situations, the agent seems to face conflicting "oughts", and even if she acts in accordance with her all-things-considered judgement, the "ought" not acted upon appears to leave a "moral remainder."[10] This moral remainder is instantiated in the phenomenology of such situations; the agent is racked with deeply negative feelings, and there is an intuition that the feelings are appropriate. A paradigm example of the type of situation that (some philosophers hold) demonstrates that unavoidable moral wrongdoing is possible comes from William Styron's novel *Sophie's Choice* (1979.) The eponymous character is forced by a Nazi prison guard to choose which of her two children's lives to save (if she refuses to choose one, then they will both die.) It is clear here that the morally correct thing for Sophie to do in this situation is to choose to save one of her children. However, it also seems as if it is appropriate for Sophie to feel deep distress. She consigns one of her children to death, and for this, very plausibly, she is morally compromised.

In the case above, Sophie's conflict is over the requirement to protect her children. In protecting one child, she fails to protect another. It is therefore a conflict over a single type of obligation. Moral conflicts of this type have been dubbed "symmetrical" conflicts. But there can also be asymmetrical conflicts, in which the conflict is between two different moral requirements.[11] A well cited example from the literature is from Sartre's *Existentialism Is a Humanism* (2007, pp. 30-33), in which a student feels that he both ought to join the French Resistance to avenge the death of his brother at the hands of the Germans during World War II, but also the he ought to remain at home and support his mother, who has already lost one son and dotes upon him. In this case, the agent appears to have two different moral obligations, only one of which can be acted upon. It seems that, whatever he does, he does wrong.

Finally, moral remainders do not only stem from situations in which the correct result of moral deliberation is disjunctive. In both cases above, the correct result of moral deliberation, all-things-considered, could be that the agent should do either A or B (O[A v B]). However, there are

---

[9] This conflict between our moral intuitions on an abstract level and when we consider concrete cases was first recognised by Adam Smith, who writes, "... [W]e may seem to be persuaded of the truth of this equitable maxim [by which he means something like RCW], when we consider it after this manner, in abstract, yet when we come to particular cases, the actual consequences which happen to proceed from any action, have a very great effect upon our sentiments concerning its merit or demerit, and almost always either enhance or diminish our sense of both." (1759, ii.iii.intro. 1–5). Philosophers who have argued for the possibility of an agent doing wrong despite doing what is all-things-considered morally best include Nagel (1972), Williams (1973), Marcus (1980), Stocker (1990), Gowans (1994, 1996) and Tessman (2015, 2017.)

[10] The term originates with Bernard Williams (1973.)

[11] Statman (1995) terms these two types of dilemmas "single value conflicts" and "incommensurability conflicts."

situations of moral conflict that leave a remainder even when there is an all-things-considered correct action. An example is the episode of the television series *The West Wing*, "Posse Comitatus," in which the US President, Josiah Bartlett, decides to assassinate an official from the fictional Middle Eastern country of Qumar because the official has been supporting terrorism. We are encouraged to see the assassination as morally justified, all-things-considered. However, it is of dubious legality, and the final scene of the episode shows Bartlett shrouded in darkness, meant to suggest the moral distress he is experiencing. The message is that even when the correct result of moral deliberation is not disjunctive, appropriate distress can remain.

Therefore, while at an abstract level RCW might seem the correct principle, particular situations such as the three outlined above give rise to an apparently conflicting intuition: that even if agents make the correct moral choice all-things-considered, they can nevertheless appropriately feel moral distress. We can state this intuition as follows:

> **In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision (MDJ).**

In this thesis, I begin from the assumption that we are justified in taking an ethical intuition *p* as providing at least some justification for *p* in the absence of defeaters.[12] We are therefore justified, in the absence of defeaters, in considering both RCW and MDJ as true. One who took these intuitions seriously could argue that the negative feelings mentioned in MDJ are best explained as picking out wrongdoing. If this is true there, then in order to maintain coherence in our moral intuitions RCW must be abandoned or somehow radically revised. If, however, there is no good reason to accept that the moral distress picked out by MDJ is indicative of actual moral wrongdoing, then we can retain confidence in RCW, especially if there is another compelling explanation of this moral distress that does not posit wrongdoing. We can also conclude that the intuition MDJ does not provide us with reason to think that UMW is possible.

A satisfactory resolution to this problem is not a matter of mere linguistic significance, for it has implications for questions of blame, the law, and punishment. For instance, if an agent can unavoidably do wrong, then plausibly she is deserving of blame—others are justified in evaluating her negatively in a moral sense, and modifying their conduct towards and relationships with her accordingly. And in some situations, the wrongdoing could be such that it was worthy of legal sanction. To return to an example I used above, if Josiah Bartlett — despite doing what is all-things-

[12] See the following chapter (1.1) for more on this.

considered morally right, nevertheless does wrong in assassinating the foreign leader, then plausibly he ought to face legal consequences for ordering the assassination. If UMW is possible, then, it seems as if it has significant consequences in the social and legal domains.

In this thesis, I contend that arguments that the negative feelings in MDJ indicate wrongdoing are unconvincing, and that these feelings therefore do not give us reason to consider unavoidable moral wrongdoing possible. I argue that the two intuitions can best be reconciled through taking the negative emotion to be agent-regret. I further argue that what agent-regret picks out is not wrongdoing, but rather a type of requirement to take responsibility for the consequences of one's (non-culpable) actions. This requirement stems from a virtue identified by Susan Wolf, which she refers to as "the nameless virtue." (2001) The thesis, then, aims to expand on Wolf's ideas in order to provide a novel account of how these two apparently conflicting intuitions may be reconciled.

In the thesis, I will largely eschew arguments against moral dilemmas (and therefore against the possibility of unavoidable moral wrongdoing) from deontic logic. While much interesting work has been done in this field,[13] such arguments generally leave the phenomenology of situations of apparent unavoidable moral wrongdoing insufficiently accounted for. As I state above (MDJ), when we consider specific situations of apparent UMW a highly compelling intuition emerges that the agent's distress is justified, and is plausibly described as guilt. Accounting for this compelling intuition—and reconciling it with RCW—is my primary concern in this thesis.

The plan of the thesis is as follows. In Chapter One, I clarify and justify the starting points for this thesis, and define terms. First, I explain the normative position from which I approach the problem. I briefly outline a position of reflective intuitionism, drawing on the approach articulated by Christopher Gowans (1994). This approach takes moral intuitions as trustworthy starting points for ethical investigation, and seeks to bring them into a state of coherence. However, I also draw on the insights of Huemer (2005, 2008) in holding that our intuitions are subject to biases, and so in addition to ensuring coherence it is also important to scrutinise them to ensure that none of these biases has infected the intuition and made it unreliable. Following that, I make clear how I understand the related terms "wrongdoing," "ought," and "blame." I also provide a brief justification for taking the terms in the way I do. Then I move on to providing prima facie justifications for the conflicting

[13] There are many arguments in deontic logic against moral dilemmas, including Conee (1982), Brink (1994) and Zimmerman (1996, Ch. 7). For arguments in favour of moral dilemmas that engage heavily with deontic logic, see Holbo (2002.) Holbo argues that deontic principles should be understood as conditionals which hold only if it is possible to avoid moral checkmate. It is pragmatically valuable for us to treat these principles as true, Holbo holds, because it will encourage us to look for a way to escape morally tricky predicaments without violating a moral rule. However, because principles of deontic logic are merely conditionally true, there may not be such a way.

intuitions that comprise the problem. Initially, I focus on justifying the intuition RCW. After that, I move on to show the plausibility of intuition MDJ. I begin by laying out the phenomenology of such situations, which consists of an emotion of moral distress coupled with an intuitive belief that the distress is justified. I then lay out the arguments of Bernard Williams and Michael Walzer for the idea that this distress is indicative of UMW, and consider several counter arguments to their view. I argue that while the counter arguments do not disprove the possibility of unavoidable moral wrongdoing, nor do Williams's and Walzer's arguments conclusively demonstrate its possibility.

In Chapter Two, I examine Gowans's "responsibilities to persons" (RTP) argument in support of the possibility of unavoidable moral wrongdoing. Gowans begins from the starting point I identified above: two conflicting intuitions about the possibility of unavoidable moral wrongdoing. He argues that we have incommensurable responsibilities to various people due to the fact that every person has unique value, and also because the nature of our relationship with each person is different. These factors, he holds, mean that there can be situations in which one has two conflicting, non-overridden moral obligations, and therefore that unavoidable moral wrongdoing is possible. To bring the intuition RCW into coherence with this, he argues that we must broaden our understanding of wrongdoing. That is, we must understand the concept also to encompass situations in which an agent performs the morally best action, all-things-considered, and yet fails to act upon non-negotiable responsibilities to persons. I argue that, while Gowans provides an intriguing solution to the problem of UMW, close scrutiny reveals that his view collapses under the weight of various implausibilities, and so fails to provide a plausible justification for the possibility of unavoidable moral wrongdoing.

In Chapter Three, I analyse Lisa Tessman's "sacred judgements" argument for unavoidable moral wrongdoing (2014, 2017). Assuming a constructivist metaethics, Tessman argues that some of our intuitively grasped moral beliefs possess a quality of "sacredness," and therefore that subjecting these sacred moral judgements to scrutiny would violate them. Because there can be situations in which these sacred judgements lead to conflicting moral requirements, unavoidable moral wrongdoing is possible, and the intuition RCW must be discarded. I argue that, even if some form of metaethical constructivism is true, Tessman's arguments for why it compels us to accept the possibility of UMW are unconvincing.

Having shown the implausibility of the two most detailed and ostensibly compelling accounts in support of the possibility of unavoidable moral wrongdoing, I turn in Chapter Four to offer a resolution of the problem that interprets the intuition MDJ in such a way that it coheres with RCW. This account utilises Wolf's (2001) concept of the "nameless virtue," developed in relation to the

problem of resultant moral luck (RML). I argue that her ideas can be adapted so as to provide a plausible solution to the problem of unavoidable moral wrongdoing. A recognition of the nameless virtue, and of the moral requirement related to it, allows for a way of acknowledging the significance of the moral remainder that exists in situations of apparent unavoidable moral wrongdoing, while exonerating the agent of wrongdoing in such situations. On this view, the moral distress the agent felt is "agent-regret" (an emotion identified by Williams in the context of the issue of moral luck), and one's requirement to take responsibility is a reflection of the general moral requirement to take responsibility for the consequences of our actions, even when one is merely causally (rather than morally) responsible for these consequences.

In Chapter Five, I expand on the argument I make in Chapter Four and address some gaps in the argument. First, I justify the contention that when an agent is causally (but not morally) responsible for harm, she nevertheless has moral obligations arising from the fact that she has caused this harm. Then, in the bulk of the chapter, I address concerns that agent-regret is irrational. I look at three existing responses to the charge of irrationality–one from Jake Wojtowicz, one from Jordan Mackenzie, and one from David Sussman–but find them implausible. I then sketch out a novel account of how agent-regret is rational, drawing on a solution to the paradox of fiction.

In Chapter Six, I address a range of objections to my proposed solution to the problem of unavoidable moral wrongdoing. These include the Generosity objection, the Observer Puzzle Objection, and the Disanalogy Objection.

In Chapter Seven, I consider at length a potential weakness with my solution to the problem of unavoidable moral wrongdoing: that it seems to require that circumstantial moral luck be possible. On Wolf's account — and my solution to the problem of unavoidable moral wrongdoing — one can accrue moral obligations through morally faultless actions (due merely to bad luck in circumstances). But if this is so, then why should bad luck in circumstances in the form of situations of unavoidable moral wrongdoing be problematic? Furthermore, if luck in *circumstances* can determine the moral obligations one has — and therefore play a role in moral evaluations — then it is unclear why it is problematic that luck in *results* should play a role in moral evaluations. But if this is the case, then Wolf's solution to the problem of RML is unmotivated. I argue that a proper understanding of the type of luck that may affect moral evaluations renders Wolf's solution — and my application of it to situations of apparent unavoidable moral wrongdoing — unproblematic.

Finally, in Chapter Eight, I analyse and clarify the nature of the requirement to take responsibility. I argue that, contra David Enoch, it is implausible to hold that, in situations in which an agent is causally but not morally responsible for bad consequences, the agent has a moral

obligation to take responsibility for these consequences, but is not, prior to this act of responsibility-taking, under the obligations that she is under subsequent to this act. I contend that, instead, an agent who non-culpably causes harm is directly under certain obligations; no initial act of responsibility-taking is required. Then, I investigate a question that has received insufficient attention in the literature: how these moral requirements or obligations are determined. Using Tamler Sommers's ideas on "negotiating blame," (2018) I make the novel argument that what is required of an agent who non-culpably causes harm is determined through a process of negotiation between the agent and the harmed party.

Overall, then, the thesis aims to contribute a novel solution to the problem of unavoidable moral wrongdoing, and in so doing offer a new justification for the rationality of agent-regret. Furthermore, it aims to strengthen the philosophical literature on moral emotions. Finally, it aims, through the "negotiating obligations" account I propose in Chapter Eight, to fill a gap in the literature on the nature of obligations that arise in situations of non-culpable harm. Such situations arise not infrequently in everyday life, and come in a variety types: apparent unavoidable moral wrongdoing, simple accidents, deliberate harm inflicted by children whom we deem too young to be moral agents. Perhaps even wrongs committed by others with whom we are intimately associated—and so for which we bear responsibility—can be dealt with under the account I propose.[14] Therefore, sketching out an account of what an agent who causes such harm owes to the harmed party or parties has the potential to have a wide range of practical applications.

[14] In a recently published article, for instance, I examine the nature of the obligations that arise as a result of historic systemic racism in Western societies. (Lane 2025)

Chapter One

Introduction

In this chapter, I aim to outline the problem of unavoidable moral wrongdoing (UMW.) I explain why each of the two conflicting intuitions I identified in the introduction seems plausible and therefore difficult to discard. It is for this reason that the problem appears such a difficult one.

In the first section, I briefly outline the normative method from which I proceed in this thesis. It is a method that draws on Christopher Gowans's "reflective intuitionism." (1994) However, I explain that I fold into this approach some insights into the circumstances which can produce bias in ethical intuitions, and therefore decrease our confidence in them. I claim that a recognition of these biases lends plausibility to the idea RCW is true.

In section two, I clarify what I mean by the term "wrongdoing," and provide a brief justification for the sense in which I use the term. I stipulate a definition of wrongdoing that equates it with acting in a way that one ought not to act, where "ought" is not necessarily to be understood as "ought all-things-considered", because this would beg the question. I further note that the "ought" in the definition is to be understood as picking out the narrow sense of "ought" identified by Henry Sidgwick (1874). I finish by commenting on the relationship between wrongdoing and blameworthiness.

Then I move on to defend the bases for the conflict I identified in the introduction. In section three, I defend the plausibility of the intuition RCW–that an action that is all-things-considered morally right cannot also be morally wrong. I begin by providing a couple of reasons for regarding it with a good degree of confidence. Then, I address some potential lines of attack on it. None of them, I argue, warrants discarding the intuition.

In section four, I defend the plausibility of taking the phenomenology of situations of apparent UMW as indicative of wrongdoing. That is, I explain why it is plausible that the justified moral distress I capture in MDJ picks out wrongdoing.[15] I do this through laying out briefly two arguments for unavoidable moral wrongdoing: Bernard Williams's argument in the essay "Ethical Consistency" (1973) and Michael Walzer's argument in "Political Action: The Problem of Dirty Hands" (1973). After

[15] Recall that in the introduction I stated MDJ as follows:

In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision

detailing these accounts, I defend them against various objections. Finally, I argue that although Williams and Walzer make a plausible case for accepting that the phenomenology picks out actual wrongdoing, it is not a conclusive one, and that further arguments are needed if we are to conclude that unavoidable moral wrongdoing is possible. I look at these alternative accounts in Chapter Two.

1.1 Reflective Intuitionism

In this section, I lay out the normative method I take in the thesis. It is not the purpose of this thesis to defend the normative viewpoint sketched out below and so I do not consider objections to the view.

I adopt a form of reflective equilibrium which takes our moral intuitions as sources of ethical knowledge. In this respect, my approach is similar to that of Gowans, whose "reflective intuitionism" I draw on.[16] I depart somewhat from Gowans however by incorporating insights from the metaethical theory of ethical intuitionism into my normative stance. I draw on the work of Huemer (2005, 2008) in doing this.

Intuitions are "initial intellectual appearances" (Huemer 2005, p. 102) that suggest that the world is a certain way. They are pre-rational; that is, one does not derive them through reasoning from other premises. They also provide prima facie justification for beliefs; if I have an intuition that *p,* I am prima facie justified in believing that *p.*[17] The justification is only prima facie, because rational scrutiny could show the intuition to be false. However, in the absence of undercutters or rebutters[18] I am justified in accepting them as true. The belief does not require additional justification apart from the intuition itself.

Ethical intuitions are intuitions that are evaluative in nature. They are ethical propositions that, when we consider them, seem to be true, prior to any reasoning occurring. Examples might include such propositions as "stealing is wrong," "honesty is a virtue," and "pleasure is better than

---

[16] Gowans's reflective intuitionism is itself based on Rawls's method of reflective equilibrium (see Rawls [1971].)

[17] Sidgwick also holds that intuitions are not beliefs, but rather provide the grounds for beliefs: "I do not mean to prejudge the question as to [an intuition's] ultimate validity, when philosophically considered; I only mean that its truth is apparently known immediately, and not as the result of reasoning. I admit the possibility that any such intuition may turn out to have an element of error… just as many apparent perceptions through the organ of vision are found to be partially illusory and misleading." (1901, p. 211) Others hold that intuitions *are* beliefs (see Audi 2004).

[18] An undercutter is a reason to doubt the veracity of an intuition, but not necessarily to abandon it. A rebutter is a confidently held intuition that directly conflicts with another intuition, giving reason to abandon the latter intuition.

pain." When one reads these statements (assuming one is competent in English and understands them), one simply "sees" them to be true. This is not to say that the propositions necessarily are true (although these three seem fairly unlikely candidates to be defeated). But it is to say that, unless we encounter reasons to call their truth into question, we are justified in holding them to be true.

These ethical intuitions are sources of non-inferential ethical knowledge, but they are not the only source of ethical knowledge. We are also able to reason from them to derive principles inferentially. For instance, I believe that Brisbane City Council ought to abolish residential zoning restrictions; that the United States government ought not to be bombing Yemen; and that I ought not to eat meat. But it seems unlikely that these beliefs derive directly from intuitions. Indeed, when I think about Brisbane City Council and residential zoning, I do not have any intuitions. My belief that its zoning policies are wrong is inferred from other beliefs, some of which are intuitively derived.

Finally, in addition to ethical intuitions there are intuitions which are relevant to ethics but which are not themselves evaluative. One set of intuitions of this type is metaethical intuitions, or intuitions about the nature of ethics. RCW is an intuition from this set. It is not itself evaluative; rather, it is *about* our evaluative judgements.

Having clarified my understanding of intuitions, I will now lay out briefly my normative method. In this thesis, I largely follow Gowans's method of "reflective intuitionism." (Gowans 1994, Ch. 2)[19] This method takes intuitions as credible but not infallible sources of moral knowledge and seeks to bring them into coherence. No principles are taken as foundational; the method seeks to bring into equilibrium substantive ethical intuitions (and the ethical principles that derive from them), metaethical intuitions, and ethically relevant non-evaluative intuitions. In cases where we identify incoherence between two or more beliefs and therefore identify a need to alter or dismiss one or more of these beliefs, we must assess the credibility of the conflicting beliefs. In doing this, the strength of the intuition is important (that is, how much it seems to be the case that *p*). However, one also needs to be aware of sources of potential bias. It might very strongly seem that *p*, but the intuition that gives rise to *p* might be the product of bias, and not pick out ethical truth. Why might an intuition be biased? I draw on Huemer (2005, pp. 140-141; 2008 pp. 381-382) in identifying the following reasons:[20]

[19] I outline this approach in Chapter Two.

[20] Interestingly, 1-3 here resemble the factors that Rawls sees as relevant to determining what counts as "considered judgments," by which he means "those judgments in which our moral capacities are most likely to be displayed without distortion." (1971/2009, p. 42)

1. **Self-interest.** Whatever is in one's particular interest can warp intuition. For instance, if I am a high-income earner, my intuition that taxation is theft might be warped by the fact that I suffer financially due to taxation.
2. **Emotion.**[21] In situations in which one is in the grip of emotion, the intuitions that arise are untrustworthy.
3. **Cultural and social traditions.** One's social and cultural background influences one's ethical intuitions (or as Sidgwick puts it, "current opinions to which familiarity has given an illusory air or self-evidence." [1874/1901, p. 212]). For instance, many in modern Western culture have the intuition that polygamy is wrong, even if all parties freely agree to the arrangement and are happy with it. [22]
4. **Evolutionary bias.** All intuitions, like all our cognitive capacities, are the products of evolution. Therefore, it is to be expected that some of these intuitions do not pick out moral truth but rather serve our own reproductive interests. Haidt's (2001) example of the intuition that it is wrong for siblings to engage in consensual intercourse using contraception is a possible example of such an intuition.

Therefore, in trying to arrive at moral truth one need not only bring his moral beliefs into a state of coherence, but also must test his intuitive beliefs for bias. Identifying that an intuition could be the product of one of these sources of bias does not show that an intuition is false; for example, an intuition could arise in a situation of high emotion and yet prove trustworthy. Nevertheless, if an intuition could be the result of one of the sources of bias listed above, then that is reason to lower one's confidence in it. It is therefore more easily defeated by other more confidently held intuitions during the process of bringing intuitions into coherence.

An important implication of this is that we are justified in considering certain types of abstract intuition to be particularly reliable. Huemer identifies three classes of ethical intuitions, namely "concrete" (or lower-level), "abstract theoretical," and "mid-level" intuitions. (2008, p. 386) While intuitions at all levels may be accurate, it is a special class of abstract theoretical intuition Huemer calls "formal intuitions" that are most trustworthy. These are abstract intuitions that "impose formal constraints on ethical theories, though they do not themselves positively or

[21] Gowans acknowledges that intuitions that we have reason to distrust arise in situations of high emotion. (1994, pp. 34-35) As he notes, this would apparently provide a reason to discount the intuition that UMW is possible. However, he goes on to argue that sometimes emotion can lead us not away from but towards moral truth, because it can draw our attention to an intuitive truth that we otherwise would not have noticed.

[22] This is not to say that one would agree with it, but rather that opposition to it would be inferentially derived from other intuitions (perhaps related to fairness.)

negatively evaluate anything [specific]." (2008, p. 386) Their reliability stems from the fact that they are unlikely to be subject to one of the sources of bias listed above. Consider the following example (from Huemer) of a formal intuition: "If *x* is better than *y* and *y* is better than *z*, then *x* is better than *z*."[23] (2008, p. 386) It is hard to see how this could be the result of emotional, cultural, or self-interested bias; and it is difficult to imagine why evolution would have produced in us false intuitions on this point. Furthermore, unlike other abstract theoretical intuitions, if we discover a counterexample, we typically do not abandon the formal intuition; rather, we are more likely to see the conflict as a paradox, and search for ways of reconciling the formal intuition with the counterexample. (Huemer 2008, p. 386)[24]

For the purposes of this investigation, the main practical consequence of incorporating an awareness of bias or error into the process of reflective intuitionism is that it makes sense to begin with a disposition to believe that unavoidable moral wrongdoing is impossible. This is because RCW is a formal intuition — it has no specific evaluative content. It is therefore unlikely to be the result of some form of bias. It is also a highly compelling intuition. Therefore, we would need very good reasons to conclude that MDJ points to the possibility of unavoidable moral wrongdoing.

To sum up, then, my own normative approach is close to Gowans's reflective intuitionism. We both begin from the assumption that intuitions are the starting points of ethical investigation. They are credible sources of ethical knowledge, but they are not infallible. If we have conflicting intuitions, then one or both of the intuitions needs to be discarded or adapted in order to iron out the conflict. However, I incorporate some insights from metaethical intuitionism into my method, a fact that will be of relevance in my analysis of Gowans's account in Chapter Two.

There is one chapter of the thesis in which I set aside the presumption that intuitions can provide us with objective moral knowledge: Chapter Three, in which I analyse the arguments of Lisa Tessman. Unlike Gowans, Tessman does not proceed from a moral realist standpoint. Rather, she adopts a position based on the antirealist constructivism of Sharon Street.[25] One way of problematising Tessman's view would be to argue that her metaethical commitments were false. However, because this would lead us substantially away from the issue of unavoidable moral wrongdoing, I will not take this tack. Furthermore, if proceeding from Tessman's constructivist commitments provides a means of reconciling the intuitions MDJ and RCW, this would be a point in

---

[23] This is also known as the Principle of the Transitivity of Value. Not all philosophers accept this principle. For arguments against it, see Rachels (1998) and Tempkin (2012.)

[24] In contrast, when we identify a counterexample to another sort of abstract theoretical intuition, we quickly abandon the intuition.

[25] See, for instance, Street (2006.)

its favour, and could warrant a reconsideration of the view I have sketched out above. Therefore, in this portion of the thesis I accept for the sake of argument Tessman's constructivism, but argue that her argument for the possibility of unavoidable moral wrongdoing collapses under the weight of its implausibilities.

1.2: Moral Wrongdoing, Oughts and Blame

In this section, I clarify how I understand three terms of central importance to this thesis: "wrongdoing," "ought," and "blame."

I shall stipulate the following definition for "wrongdoing":

> Wrongdoing occurs when an agent acts in a way in which she ought not to act.[26]

This seems a fairly uncontroversial understanding of wrongdoing. Importantly, it does not beg the question in favour of one side of the debate over whether unavoidable moral wrongdoing is possible, as I explain below in my discussion of the term "ought."

W. D. Ross (1931/2002) held that "oughts" were all-things-considered duties. At this point in the thesis, I do not take a position on whether "ought" in the above definition picks out "ought all-things-considered" or not. As the debate is over whether an agent can do wrong despite doing what is all-things-considered right, such a definition would beg the question — if an agent only does wrong when she does what, all-things-considered, she ought not to do, then UMW must necessarily be impossible. Instead, then, I remain uncommitted as to whether "ought" is all-things-considered, or whether it picks out Peter Vranas's understanding of the term: "that an agent ought to do something [is] the claim that the agent has an objective, pro tanto obligation to do the thing." (2007, p. 169) [27] Furthermore, "ought" is obviously meant in a moral sense. The word "ought" can also be used

[26] Tessman prefers "failure" to "wrongdoing" because she takes "wrongdoing" to refer to the violation of an all-things-considered duty, and not just the violation of a moral requirement. I follow Gowans' use of the term "wrongdoing" because, as Tessman notes, it is the term commonly used in the literature (2014, p. 15 footnote). Indeed, despite her comments, Tessman uses it herself at times (see Tessman 2014, Ch. 1-2.)

[27] Stocker (1990, Ch. 1) uses "ought" in this way--to pick out apparent moral requirements. He follows Williams in arguing that moral oughts are like desires in the sense that when they conflict the "ought" not acted upon is not eliminated. He justifies this by arguing from analogy to costs. Everything we do has a cost, and even if we deem that the benefit outweighs the costs, the costs are not eliminated. They are therefore counted twice–"once in taking them as costs and once in holding that despite their being costs, what they are costs for is justified." (1990, p. 15) Double-counting is therefore conceptually unproblematic, and as not acting on an "ought" is a type of cost, it is also unproblematic to hold that "oughts" can be double-counted (and so the one not acted upon is not eliminated in a case of conflict.)

prudentially (for instance, when one says, "The weather is nice today, so I ought to go to the beach"), but this sense is clearly not the one meant in the above definition. Finally, not only is "ought" meant in a moral sense, but it is meant in a particular moral sense. Sigdwick notes that at times we use "ought" in what he terms a "wider sense." This wider sense is the one that is picked out when we state that the "laws of [our] country ought to be other than they are." (1901, p. 33) In other words, it expresses a morally desirable state of affairs, but does not imply that any particular agent has an obligation to bring about the state of affairs.[28] The sense in which I use "ought" in this thesis is the narrower sense that picks out agential obligation.

Gowans and Tessman both use "requirements" rather than "ought" (or the related noun, "obligation"[29]) in this context. Gowans chooses "moral requirements" because he feels "ought" has tended to be used to refer to the result of an all-things-considered deliberation (1994, p. 48). Tessman explicitly equates "moral requirements" with "oughts" (2014, p. 22). Nothing hangs on these differences in terminology, and when I write that an agent "ought" to do *x* it can be considered to have the same meaning as the proposition, "the agent is morally required to do *x*." When I intend "ought" to pick out the result of moral deliberation, I use "ought all-things-considered."

Finally, let me turn to "blame." I proceed from the tentative assumption that an agent is blameworthy if and only if the agent has done wrong. Furthermore, I assume that if an agent does wrong, then an agent is necessarily blameworthy (in proportion to her degree of wrongdoing.) In contrast, both Gowans and Tessman suggest that it is possible that one can do wrong without being blameworthy (Gowans 1994, pp. 136-138; Tessman 2014, p. 16); and so in order to avoid begging the question against their views I do not assume this necessary connection between wrongdoing and blameworthiness in assessing their arguments. However, thereafter (that is, from Chapter Four onwards), I assume that it is the case that an agent is blameworthy in proportion to her degree of wrongdoing.[30]

---

[28] Others who have recognised the different sense of "ought" include Vranas (2007 p. 175) and Wedgewood (2009 p. 503-504.)

[29] I follow Singer (1973) and Talbot (2016) in assuming that the proposition "S ought to do *x*" is identical to the proposition "S has an obligation to do *x*."

[30] For arguments against this view, see McKenna (2012) and Buckwalter (2020). Fritz (2014) provides (to me) a compelling response to views such as McKenna, in which he distinguishes between a) an agent S being deserving of blame and b) the appropriateness of a particular agent blaming S. Any agent who does wrong is deserving of blame in proportion to his degree of wrongdoing, but we will not necessarily blame him unless we stand in a certain relation to his act of wrongdoing.

1.3: The Plausibility of RCW

In the next two sections, I aim to argue for the prima facie plausibility of the two conflicting intuitions, RCW and MDJ. I do not contend that the arguments in these sections demonstrate categorically that neither RCW or MDJ should be discarded; I merely hold that these arguments demonstrate their plausibility and that neither should be summarily dismissed in order to resolve the problem of unavoidable moral wrongdoing. Either one may, upon closer scrutiny, prove to be worthy of discarding; but these sections demonstrate that much more detailed arguments will be required to show it to be so.

In this section, I focus on the plausibility of RCW. One reason we have for considering it plausible is that, as a formal intuition, RCW is — in accordance with arguments I made above (see 1.1) — less likely to be subject to bias. Then, I argue for RCW through showing how it accords with another very plausible intuitive belief about moral luck. Subsequently, I address some potential arguments against the intuition and argue that they are not powerful enough to warrant dismissing it.

1.3.1: RCW as a Formal Intuition

As I noted above (see 1.1), a principle of my normative method is that formal intuitions —intuitions that do not contain evaluative content themselves, but "impose formal constraints on ethical theories" (Huemer 2008, p. 386) — should be afforded a greater presumption of truth than lower-level intuitions. RCW is a formal intuition. This means that the chance that it is the product of one of the sources of bias I identified earlier is low. It seems highly improbable that RCW could be the result of emotional bias. Indeed, when I have the intuition, I feel very little emotion, if any. It is also unlikely that it could be the result of self-interest or cultural bias. That RCW is true does not appear to benefit me in particular; and it is not obvious that it could be the product of any particular cultural beliefs. Finally, it is hard to imagine how it could provide a selective advantage to believe RCW. Therefore, we may conclude that it is unlikely that the intuition is the product of any form of bias.

1.3.2: Moral Luck

A plausible principle in ethics is that the degree to which one acts morally cannot depend on factors outside one's control. This is known as the control principle (CP). In other words, it seems counter-

intuitive that there can be "moral luck".[31] If one could, in doing what is all-things-considered morally right, commit wrongdoing (in other words, if RCW is false), then the control principle would have to be rejected.[32] This is because if (as I assume), it is correct that there are no true dilemmas (either one choice outweighs the other or there is a disjunctive requirement), then in each situation there is a right way to act. However, if in making that right choice an agent is also somehow acting wrongly (in other words, if RCW is not always true) then it cannot be true that agents are only morally judged for what is under their control (as in such a situation the wrongdoing would be out of the agent's control).

Therefore, if we want to retain our belief in CP, we cannot abandon RCW. As CP is a formal intuition, it is unlikely that it is subject to one of the biases I identified above (see 1.1). We therefore have no reason to withdraw confidence from it on those grounds. This obviously does not prove that RCW is correct; perhaps there are arguments in favour of UMW that are so convincing that we must decide we need to reinterpret, modify or abandon both RCW and CP. However, that RCW coheres with another very appealing intuition, CP, gives us at least some reason to think that we do not go awry in accepting it as true.

1.3.3: Objection from the Falsity of "Ought" Implies "Can"

One argument against RCW might involve the supposed falsity of the principle "ought implies can" (OiC). OiC states that, if a person ought to do something, then she can do that thing. In conjunction with another principle of deontic logic, the "agglomeration principle", it has been used to argue that UMW is impossible.[33] If OiC is false, then RCW must also be false. This is because, if there is a situation where one ought to do *x*, but one is unable to do *x*, then even if one performs the morally correct action out of the set of actions that he can do, then he has nevertheless done wrong. In other words, he has done wrong in doing right and RCW is false.[34]

---

[31] Nagel (1979) identifies four types of moral luck: resultant, circumstantial, causal, and constitutive. Recent philosophical attention to the problem of moral luck stems from this paper as well as one by Bernard Williams (1981.)

[32] The connection between situations of apparent unavoidable moral wrongdoing and moral luck is under-appreciated in the literature. Van Zyl (2007) identifies such situations as examples of circumstantial moral luck.

[33] See, for instance, Conee (1982.)

[34] This would not be the case if one held that there could be situations in which there was no right decision (*i.e.* that all possible actions were wrong.) As I said in the introduction, I am assuming that there are no such situations.

The intuition OiC seems to me very plausible. However, some recent experimental work has called into question whether this view of its plausibility is widely held. Chituc et al. (2016) for instance ran a series of experiments in which participants read brief stories featuring agents who were unable (for either a low-fault or high-fault reason) to perform a task such as meeting a friend at an appointed time Participants were then asked to rate on a 100 point scale the extent to which the agent could perform the action; the degree to which he ought to perform the action; and the degree to which he was to blame for not performing the action. They found that the ratings on the scale given for "ought" did not correlate with the ratings for "can." In one experiment (a "high-blame" scenario in which the agent is responsible for his inability to fulfill a promise), 60% of respondents said that the agent nevertheless still ought to fulfill it. Rather, Chituc et al. found that, "ought" judgements correlated with "blame" judgements, with respondents tending to hold that an agent ought to perform impossible actions when he was to blame for not being able to perform them, but not when his inability to perform them was out of his control.[35] This suggests that, despite the fact that some philosophers think that OiC is intuitive, this intuition is not widely held. Rather–as Paul Henne et al put it–they "seem to conflate 'ought' with blame.'" (2016, p. 286)

One response to this is to attribute performance errors to participants of studies such as that of Chituc et al. Shaun Nichols and Joshua Knobe, writing on the subject of free will, suggest that

> people's responses to questions about moral responsibility can vary dramatically depending on the way in which the question is formulated. When asked questions that call for a more abstract, theoretical sort of cognition, people give overwhelmingly incompatibilist answers. But when asked questions that trigger emotions, their answers become far more compatibilist.
>
> (2007, p. 664)

Later, they note that

> people give compatibilist responses under certain [affect-laden] circumstances, [but it is possible that] there is not any real sense in which people can be said to hold a compatibilist view of moral responsibility. Instead … the compatibilist responses we find in our concrete conditions are to be understood in terms of performance errors brought about by affective reactions. In the abstract condition, people's underlying theory is revealed for what it is—incompatibilist.
>
> (2007, p. 671)

If this is correct, then it could be argued that the responses of participants in Chituc et al. are subject to similar performance errors. This is because participants in Chituc et al. were also presented with

[35] See Buckwalter and Turri (2015) and Mizrahi (2015) have reported similar results with respect to the relationship between "ought" and "can."

vignettes designed to test responses to "concrete" situations that might evoke affective responses. It is possible that, had participants instead been asked more abstract questions about the relationship between "ought" and "can," a different result would have been reached.

But even if we accept the results of Chituc et al. problematising OiC, they might nevertheless offer support for the thesis that unavoidable moral wrongdoing is impossible. This is because their findings suggest that people only blame agents for not performing impossible actions when the agents themselves are responsible for their inability to perform them. In most cases of unavoidable moral wrongdoing (including the paradigm cases drawn on in phenomenological arguments) this is not the case—Sophie, for instance, is hardly blameworthy for being imprisoned by the Nazis and accosted by a sadistic guard. Therefore, if an agent only ought to do the impossible in cases in which he is responsible for this impossibility, we have good reason for dismissing many of the arguments philosophers have made in favour of the possibility of unavoidable moral wrongdoing (including those of Gowans and Tessman, which I address at length in Chapters Two and Three.)

A stronger argument against OiC has attracted surprisingly little attention in the literature.[36] It comes from Todd Bernard Weber (2002.) In brief, Weber argues that to use OiC as part of a logical proof to demonstrate that unavoidable moral wrongdoing is impossible is to beg the question. This is because OiC is simply another way of stating that UMW ("inescapable wrongdoing" in Weber's parlance) is not possible. Weber cites the following passage from David Brink to show this:

> It seems … intuitively compelling … It may seem too harsh to require of people actions that they cannot perform; the voluntarist insists that it be up to us whether to be moral. Nonetheless some people claim that there are counterexamples to the voluntarist principle; they claim that there are impossible obligations.
>
> (Brink 1994, cited in Weber 2002, p. 465)

In other words, we have the intuition that "it is up to us whether to be moral," and so find OiC plausible based on the appeal of that intuition. (2002, pp. 464-465) However, he argues, "it is up to us whether to be moral" is simply another way of stating that inescapable moral wrongdoing is impossible. If this is the case, then deontic logical arguments using OiC to show that UMW is impossible beg the question.

One could respond that Weber offers insufficient proof for the view that OiC is derived inferentially from the intuition that "it's up to us whether to be moral." The fact that Brink, in arguing for the plausibility of OiC, suggests that it is the case does not, after all, prove that it is so. Perhaps the

[36] An exception to this is Stephen De Wijze's recent paper arguing in favour of "dirty hands," where he endorses Weber's view. (2022, pp. 201-202)

two intuitions emerge separately but are highly similar in their content. Therefore, at best Weber's argument presents an intriguing possibility.

More problematic however is the view that the proposition "it's up to us whether to be moral" is the equivalent of the proposition "Unavoidable moral wrongdoing is impossible." The fact that these two propositions are not semantically identical can be seen by the fact that the following syllogism is valid:

P1. It's up to us to be moral.
P2. If unavoidable moral wrongdoing is possible, then it is not up to us to be moral.
C1. Therefore, unavoidable moral wrongdoing is not possible.

If C1 was identical to P1, then P2 would be redundant. But P2 is not redundant, and so Weber's argument cannot be correct on this point.

Even so, "it's up to us whether to be moral" *is* arguably simply a (basic) expression of the control principle; and so if Weber is correct, then the plausibility of OiC is not a distinct piece of evidence in favour of the plausibility of RCW, but rather a restatement of the argument from the plausibility of the control principle. Therefore, it *would be* illegitimate to argue that OiC provides a substantial piece of evidence in support of RCW. But as I have not made such an argument, his view — even if correct — does not appear to jeopardise the confidence we can repose in RCW.

1.3.4: Objection from Common Linguistic Usage

Another objection to accepting RCW as pointing to truth could be that the way people use language belies the idea that an action cannot be both right and wrong simultaneously. Consider a situation in which two married people are having an affair. "It's so wrong but it feels so right," one of them might say to the other.[37] In this case, it might appear, the affair is both wrong (in that it breaks an obligation of fidelity each partner has to his or her spouse) and right (in that it allows both partners to flourish or be true to themselves.) Therefore, the objection goes, there is nothing conceptually problematic about something being both wrong and right at the same time.

But this objection also fails to go through. In the above example, it seems highly plausible that while "wrong" picks out moral wrong, the same cannot be said for right. "Right" in the sentence

[37] An example of this ubiquitous trope can be found in the Carly Rae Jepson song "Wrong Feels So Right." (2012)

instead seems to pick out prudential rightness or desire satisfaction — it is all-things-considered good for the two participants in the affair that it should continue, or it satisfies certain of their desires for it to occur. However, it is not morally right. If this is correct, then the fact that this is a common trope does nothing to detract from the plausibility of the intuition RCW.

1.3.5: Remaining Questions for RCW

I have argued that, outside of the fact that particular instances (real or hypothetical) can produce the intuition that some form of moral distress is appropriate, there are no obvious reasons to dismiss RCW. It is a formal intuition, and so is unlikely to be untrustworthy; and it coheres with another plausible principle, the control principle. Furthermore, while there are some arguments against OiC, these are insufficient to call RCW seriously into question. Finally, while it is possible that there is some additional sense of "wrong" which could render RCW false, this is not obviously the case, and additional argument is needed to demonstrate this.

I now proceed to establish the plausibility of MDJ.

1.4: The Plausibility of MDJ

In the previous sections, I argued that there were no obvious reasons to consider the intuition RCW implausible. In the rest of this chapter, I show that the same is true for MDJ–the intuition that the emotion of moral distress experienced in situations of apparent unavoidable moral wrongdoing is appropriate. First, I lay out what I take to be the role of the emotion of moral distress in situations of apparent unavoidable moral wrongdoing. In order to do this, I lay out how Williams (1981) provides a plausible argument that the phenomenology of these situations picks out actual wrongdoing. As with the arguments in the previous section, I do not claim that the arguments I put forth here are dispositive. In this section, I summarise Williams's argument, highlighting its insightful recognition of the "moral remainder" in cases of apparent UMW. I then discuss various criticisms of Williams's account, and argue that most fail to disprove it. I finish by agreeing with a point made by Gowans: that although criticisms of Williams's account do no not show it to be false, Williams provides no positive argument for why we should take the existence of moral remainder to be indicative of wrongdoing.

1.4.1: The Emotions in Situations of Apparent Unavoidable Moral Wrongdoing

In order to show why I consider MDJ plausible, it will help if I first clarify what I consider to be the significance of emotions in situations of apparent unavoidable moral wrongdoing. I will first summarise the phenomenology of such situations; and then I will explain the significance of emotion in these situations.

As I have noted, the phenomenology of situations of UMW is characterised by appropriate distress on the part of the agent. There are two important elements to this phenomenology: its intentionality and its appropriacy. I will briefly address both elements. First, the intentionality of this distress is at the agent herself. That is, the emotion is self-directed. It is not that the agent simply has negative feelings about the state of the world or about some third party, or that she is morally pained about some event to which she has no connection, such as human rights violations in a foreign nation. It is that she herself is (at least in part) the object of the emotion. Second, the agent (and third-party observers) consider the distress to be appropriate. Neither she nor anyone else who is aware of all the facts of the situation think that the distress is unjustified in the way that one might consider fear unjustified upon discovering that the snake she is afraid of is not venomous, or is merely a convincing plastic replica.

What do I mean when I say that an emotion is justified (or appropriate, or rational — I use the terms interchangeably)? An emotion is appropriate when it picks out the evaluative concept to which it is connected.[38] For instance, for fear to be justified is for the fear to pick out a dangerous situation; and for moral distress to be justified is for it to pick out a situation that is morally distressing. What is the source of this justification? Some hold that the emotion itself provides at least prima facie justification. Elgin argues for instance that emotions are "deliverances [that] are in some respects like perceptual deliverances" (2008, p. 48), and that they are therefore "indicators of how things stand" which provide "some initial degree of tenability."[39] (2008, p. 40) That is, to feel fear provides at least some evidence that the object of the emotion is dangerous.[40] Others argue that emotions provide no epistemic justification. Brady illustrates this view with reference to an encounter with a dangerous animal. When one sees that "a large dog has sharp teeth, a short temper, is off its lead, and is advancing rapidly towards me," the fear one feels towards it is not a reason to consider it dangerous. Rather, one considers it dangerous because of its sharp teeth, the fact that it is snarling and drooling and off its leash, and so forth. (2011, pp. 142-143) On this view, emotions are

[38] Although in Chapter Five I argue that there is another way in which an emotion may be rational.
[39] For other views along these lines, see Goldie (2000); Doring (2003); and Tappolet (2011, 2016.)
[40] Audi (2013, Ch. 6) seems to hold a similar view.

justified or appropriate for reasons independent of the emotions themselves. The fear I feel is not an *additional reason* to consider the animal dangerous, but a *product* of or response to reasons for considering it dangerous. The emotion might play the role of drawing attention to a situation in order to prompt a search for reasons; but it does not have any independent epistemic value.

It seems implausible to suggest that emotions can never play any justificatory role.[41] However, in this thesis I do not assume that, in cases of apparent unavoidable moral wrongdoing, the emotion of distress plays any role in justifying itself. Rather, I contend that the emotion is *prima facie* appropriate because of the intuition MDJ. This intuition tells us that the situation is morally distressing, and therefore that it is appropriate for the agent to feel moral distress.

One might ask why we could not simply examine the moral distress, determine what exactly it is, and from this draw a conclusion about whether the agent had committed wrongdoing. For example, if the moral distress was guilt, then we could conclude the agent had engaged in wrongdoing (on the assumption that when an agent feels guilt, and we judge the guilt is appropriate, he has done something wrong.) However, this in my view misunderstands the nature of emotions. As Robinson has argued, an emotion is a process, consisting of both "affective and cognitive evaluations." (2004, pp. 72-73) The affective part of the emotion is automatic and immediate, and paints in broad strokes; these evaluations may be no more specific than "I don't like this" or "eww." As Robinson puts it, "these evaluations cannot make fine discriminations." (2004, p. 72) On Robinson's view, what allows us to distinguish among the multitude of emotions for which we have names is the cognitive activity which follows this affective evaluation.[42] This cognitive activity is not separate from the emotion, but is rather a part of the emotion, because the affective appraisal can be "modified by subsequent cognitive appraisals and reappraisals." (Robinson 2005, p. 78) In other words, emotion is a process that begins with an affective evaluation, which triggers a cognitive evaluation. This cognitive evaluation may then feed back into the affective evaluation and modify it, which may then trigger further cognitive evaluation.[43] In the case of a woman who feels intense negativity towards an attractive young female

---

[41] For instance, when one listens to Beethoven's "Sonata Pathetique," it seems implausible to suggest that one needs to search for reasons other than the emotion itself to justify the belief that the piece is sad. Brady suggests that, in such cases, while we do not generally search for additional reasons, they are in principle available (in the case of "Sonata Pathetique," perhaps he would point to the fact that it is in a minor key) and this sets emotions apart from sense perception. However, this is unconvincing. When we look at an object and it seems red, there are also in principle other reasons available to justify it (concerning the wavelengths of the light as it reflects off the object and son). But it would be strange to suggest that our seeing an object as red did not provide justification for the fact that it was red.

[42] See also Greenspan (2004, pp. 221-222.)

[43] Robinson refers to her account as non-cognitive, but, as is apparent from my summary, it is not clear why this is. Given that emotion is a process which involves cognitive evaluation, it seems more accurate to say that emotion *begins* non-cognitively, but quickly becomes cognitive.

colleague of her husband's, then, the affective evaluation is something like, "I don't like." It is only on the basis of the subsequent cognitive evaluation that she might come to recognise the emotion as jealousy.[44]

However, as Robinson notes, there are times at which the appropriate cognitive evaluation is not clear:

> When I am distraught [that is, experiencing a negative affect] about my husband's abandoning me, I do not know exactly what emotions I am experiencing… I am in a state of arousal, but my state is ambiguous and I cannot tell exactly why I am responding as I am.
>
> (2005, p. 86)

In situations such as these, where it is not immediately clear why one is experiencing a particular affect, it may only be upon extended cognitive processing that the agent determines what the appropriate cognitive evaluation is.

The relevance of this to the agent's moral distress in situations of apparent unavoidable moral wrongdoing is that one cannot simply know from the affect in these situations what the emotion is; that is, the affective appraisal of distress cannot be immediately identified as guilt. On the account I have sketched out above, guilt is an affect or bodily feeling of distress plus a cognitive component—which may precede or follow the affective component—that one has done wrong. But whether this is the appropriate cognitive component is unclear in cases of apparent UMW. The bodily feeling or affect is similar or identical to that one experiences when one feels guilt; but the process of refining the broad strokes of the initial evaluation does not yield a clear result. The refinement process seems to lead to an evaluation along the lines of, "I wish I had not done this." But it is unclear that this necessarily picks out wrongdoing, and so there is uncertainty over whether one has actually done wrong in these cases.

To sum up, the phenomenology of situations of apparent UMW consists in the agent experiencing an emotion of distress which is self-directed and which is appropriate because it is supported by the intuition MDJ. However, this does not necessarily mean that in such a situation the agent has committed wrongdoing. Perhaps there is some other explanation for why these justified self-directed feelings are appropriate. But because self-directed moral distress typically picks out

---

[44] A view that is along the same lines is that espoused by Whiting (2011). On Whiting's view, an emotion is simply the affect or bodily feeling. What gives us the sense that (some) emotions have intentionality (or are, as Whiting puts it, "object-directed") is a separate cognitive component or thought that precedes the emotion. Thus, the emotion *fear-of-tiger* is actually a combination of an initial thought– "the tiger is dangerous"--followed by the affect of fear (an edgy bodily sensation). For this reason, Whiting terms his view the "composite view" of emotions. Whiting's view therefore differs from Robinson's in terms of where the cognitive component is situated. On Robinson's view it comes afterwards, whereas on Whiting's it comes first.

wrongdoing, it seems plausible that it does. In the following section, I lay out Williams's argument to provide further support for this view.

1.4.2: Williams's and Walzer's Arguments for the Possibility of Unavoidable Moral Wrongdoing

In the remainder of this chapter, I use the arguments of Bernard Williams and Michael Walzer to show how the view that MDJ picks out wrongdoing is plausible. I briefly lay out their accounts; consider some objections to them; and hold that none is decisive. However, while none is decisive, nor do Williams or Walzer provide sufficient proof that the moral distress picks out wrongdoing to justify the dismissal or drastic modification of RCW.

Williams's account aims to dispel the intuition I have dubbed RCW. Williams argues that moral beliefs (or moral "oughts") are like desires rather than beliefs. Let's say I have belief *a* that there will be a thunderstorm tomorrow and also belief *b* that it will be sunny for the entire day tomorrow. In this case, *a* and *b* conflict, and upon recognition of this I will critically examine them to determine which is true and then discard the false one. In other words, when beliefs conflict, one of the beliefs must be eliminated. Desires, however, do not work like this. If I desire it to be sunny all day tomorrow because I want to go to the beach, but also desire there to be a thunderstorm tomorrow because I enjoy the feeling of sitting on my (covered) patio watching rain pour down, I may decide that my desire for it to be sunny outranks my desire for there to be a thunderstorm, but the desire for there to be a thunderstorm is not eliminated. It remains on the scene. Williams argues that moral oughts are like desires rather than beliefs, in that in situations of conflict the defeated moral ought — the moral ought not acted upon — is not eliminated.[45] Rather, it leaves a remainder, a state of moral distress which "acknowledges the presence of both the two oughts [the ought acted upon and the ought not acted upon]." (Williams 1973, p. 172) Williams holds that this remainder is evidence of wrongdoing:

> If a man in general thinks that he ought not to do a certain thing, and is distressed by the thought of doing that thing; then if he does it, and is distressed at what he has done, this distress will probably have the shape of his thinking that in doing that thing, he has done something that he ought not to have done.
>
> (1973, p. 174)

[45] See De Wijze (1996) for further arguments along these lines.

Here, Williams holds that the relationship between *(a)* the thought that one ought not to do X and *(b)* the distress caused by one's doing X is best explained by the idea that he has done wrong: "*(b)* [is] one expression of *(a),* and *(a)* one root of *(b)*." (1973, p. 174) And nor does Williams contend that it is the case that an agent who does X and then experiences distress is unjustified in experiencing it. With reference to the case of Agamemnon, who must sacrifice his daughter Iphigenia in order to placate the goddess Artemis and for his ships to continue their voyage to Troy, Williams notes that even if we accept that Agamemnon makes the better choice "it would seem a glib moralist who said, as some sort of criticism [of Agamemnon], that he must be irrational to lie awake at night, having killed his daughter." (1965, p. 111) In other words, third party observers also consider the distress justified, and so the sense of justification cannot be dismissed as an error on the part of the agent. This, Williams argues, provides evidence that in situations of apparent UMW a person does wrong even if she makes the correct moral judgement.

Williams further argues that the phenomenology of situations of apparent UMW is good evidence for actual wrongdoing because to consider the two conflicting oughts mere *prima facie* oughts does not explain the distress an agent feels. Ross (1931/2002) holds for instance that all obligations are *prima facie* obligations, and that when they conflict one must use one's judgement to determine which prevails in that particular situation. However, Williams notes that if this is true it is not clear how a *prima facie* obligation, having been defeated, can continue to exert a hold on one that manifests itself in negative feelings. If the obligation is merely *prima facie,* then the agent should feel "relief" or something similar at having escaped a moral mistake (1973, p. 176). The nature of the remainder therefore shows that the obligation is not a *prima facie* one.

Although Walzer's argument is made largely with reference to UMW among political actors,[46] it is broadly similar to that of Williams. Walzer asks us to imagine a situation in which a recently elected politician is dealing with

> a terrorist campaign, and the first decision the new leader faces is this: he is asked to authorize the torture of a captured rebel leader who knows or probably knows the location of a number of bombs hidden in apartment buildings around the city, set to go off within the next twenty-four hours. He orders the man tortured, convinced that he must do so for the sake of the people who might otherwise die in the explosions even though he believes that torture is wrong, indeed abominable, not just sometimes, but always.
>
> (1973, pp. 166-167)

[46] Walzer coined the phrase "dirty hands" in this essay to characterise situations in which a politician or government official has to commit an apparently wrong act in order to satisfy his obligations to the larger polity he represents.

In this case, Walzer holds, the politician does the right thing by ordering the rebel leader tortured. However, in doing what is all-things-considered morally right, "he committed a moral crime and he accepted a moral burden. Now he is a guilty man." (1973, p. 167) That this guilt is appropriate is made clear by a thought experiment involving a member of a firing squad. Walzer notes that members of a firing squad are not told whether their rifles contain blanks or live rounds, the purpose of which is to allow each member to avoid or lessen any feeling of moral distress at having caused the death of another human being. This trick works, Walzer argues, when the members of the firing squad believe that the man being executed has committed a crime and deserves his punishment. However, it would not work if they knew he was innocent, and was being executed for "some overriding political or moral reason." (1973, p. 174) In this case (like the first case), the executioner believes that the killing is all-things-considered morally right. However, even if he did not know whether his rifle contained a live round or not, he would still feel deep moral distress over his actions. As Walzer puts it, "[o]ur guilt feelings can be tricked away when they are isolated from our moral beliefs, as in the first case, but not when they are allied with them, as in the second." (1973, p. 174)

Both Williams and Walzer argue, then, that the intuition MDJ is plausible, and that it picks out moral wrongdoing. Below, I consider two objections.

1.4.3: Objection from Unjustified Emotions

One argument that has been made against accounts such as Williams's and Walzer's is that emotions are commonly inappropriate and that situations of apparent unavoidable moral wrongdoing are simply another instance of this phenomenon. For instance, Kai Nielsen argues that people commonly experience feelings of moral distress; but that "to feel guilty is not necessarily to be guilty. Plenty of people feel guilty without being guilty… ." (2000, p. 140) As the feelings of distress are unwarranted, there is no problem of dirty hands (that is, unavoidable moral wrongdoing is not possible.)[47]

However, Nielsen fails to take into account that, on the view of Williams and Walzer, the agent's moral distress is justified. As I noted above, Williams holds that agents in situations of apparent UMW experience distress about what they have done, and that both the agent and third party observers consider this distress justified. Walzer's example of the member of the firing squad

[47] At least, there is no problem of dirty hands in the sense that Walzerian sense. Nielsen argues that there *is* problem of dirty hands, but that it concerns the question of "when, if ever, are we justified in using what would in normal circumstances clearly be a morally impermissible means to achieve what is clearly a morally demanding end?" (2000, p. 153)

whose moral distress cannot be assuaged makes a similar point. In other words, according to Williams and Walzer agents in situations of apparent unavoidable moral wrongdoing consider the distress justified, and justified on the basis of wrongdoing on their part. If Nielsen were correct, then the agent would feel distress, but she would know on some level that her distress was inappropriate (or, at least, third party observers would know this.) But she does not know this. On the contrary, she thinks it is appropriate. Therefore, Nielsen's arguments are unconvincing.

1.4.4: McConnell's Objection

Terrance McConnell's objection concerns not whether the emotion is justified, but the nature of the emotion. McConnell holds that it is plausible that the distress an agent experiences in a situation of apparent UMW does not pick out agential wrongdoing. He argues that we have better grounds for taking it to be regret at "a bad state of affairs" that has been brought into existence. McConnell holds that when an agent has done wrong, the experiential component (the negative feeling) is accompanied by a cognitive component (a belief that one has done wrong.) In cases of apparent unavoidable moral wrongdoing, it is plausible that this cognitive component is absent and that the emotion is simply regret. (1996, p. 38) If it is regret, it allows us to retain appealing principles of deontic logic, and so "unless there are special reasons for taking this feeling to be remorse, it seems that we are justified in taking it to be regret." (1978, p. 278)

Unfortunately, for McConnell, there is a reason for taking it to be remorse: the intentionality of the emotion. McConnell's argument is that the emotion of distress can be accounted for by the "bad state of affairs" that the agent's action has produced. Therefore, the correct cognitive component accompanying the affect must be, on his view, something like, "it would be good if this had not happened." But this is not the intentionality of the emotion an agent experiences in a situation of apparent unavoidable moral wrongdoing. As I noted above (see 1.4.1), in cases of unavoidable moral wrongdoing, the negative affect is (at least in part) self-directed. It is not simply some negative feeling directed towards some event; it is a negative feeling or distressing emotion that is directed at oneself. As McConnell's account states that the distress simply picks out a bad state of affairs, it is unclear how it can explain this intentionality.

1.4.5: Objections from Subjectivity

A similar objection comes from Conee (1982) and concerns the subjective appropriateness of the feelings of distress. He agrees that the distress is justified and that its intentionality is towards the agent herself. However, he argues that the justification is subjectively correct, while being incorrect from an objective standpoint:

> Feeling guilty is subjectively appropriate when the belief that one has failed which prompts the feeling fits one's moral principles. If your convictions include that every debt morally must be repaid, it is appropriate to your morality for you to feel guilty about defaulting. When someone does what is morally best while neglecting something his morality requires, his feeling guilty is therefore appropriate only because it is called for by morality as he sees it. It does not fit the facts. This sort of appropriate guilt does not imply that a moral mistake has been made. So an opponent of moral dilemmas can consistently hold that feeling guilty about a morally superior act is clearly appropriate at times, but in this subjective way, not in light of any omitted actual moral obligation.
>
> (1982, p. 91-92)

Here, Conee seems to hold that, if an agent feels guilty despite acting in accordance with his all-things-considered moral judgement, it could be subjectively appropriate if the agent believes that he has violated an absolute moral principle.[48] It is subjectively appropriate because guilt is appropriate when one has done wrong, and the agent thinks he has done wrong. Because he has not done wrong, however, his emotion is objectively inappropriate. Applying this to Williams's argument, Conee could hold that the agent does feel guilt, but it is only subjectively appropriate and does not pick out actual wrongdoing.

One response to this argument is to note that it is not only the agent herself who considers her feelings of moral distress appropriate — third parties concur. For instance, if after executing an innocent man for overriding moral reasons, the executioner were simply to shrug off his actions and proclaim that he felt no self-directed distress at all, we would consider this deeply inappropriate. As Walzer puts it, the executioner "will know that he is guilty (and we will know it too.)" Therefore, the intuition MDJ cannot be explained away as an inaccurate — if inevitable — response on the part of the agent to situations of moral conflict.

A second response to Conee is that he simply begs the question.[49] Conee holds that the agent's guilt can only be subjectively appropriate because the agent has, all-things-considered, done the

[48] Foot (2002) makes a similar point. Sidgwick also speaks of acts being "subjectively wrong" when the agent believes them to be wrong even though they are "objectively right" or "materially" right. (1901, p. 207)
[49] Gowans makes a version of this point. (1994, p. 107)

morally right thing. If it is only subjectively appropriate, then it is not indicative of wrongdoing. But Coney can only arrive at this conclusion that it is merely subjectively and not objectively appropriate by first asserting the premise that the agent has not done wrong. As this is also the conclusion, his argument is not valid. At best, Conee's assertion (that the agent mistakenly considers himself to have done wrong) provides an alternative interpretation of the negative feeling that an agent has in a case of apparent UMW, but his argument does not show this.

1.4.6: Williams's and Walzer's Argument as Non-Decisive

We have seen that objections to the view of Williams and Walzer are unconvincing. However, while arguments for its falsity do not go through, nor do Williams or Walzer do enough to prove decisively that UMW is possible. In this section, I explain why.

Laid out formally, Williams's and Walzer's argument can be put as follows:

P1. If an agent experiences moral distress and believes that it is justified, then he has committed wrongdoing.
P2. In situations of apparent UMW, the agent typically feels moral distress and believes it is justified.
C1. Therefore, the agent has committed wrongdoing.

As the argument is valid, to render the conclusion false either P1 or P2 needs to be questioned. P2 seems clearly true, but P1 is more problematic. This is because it is not clear that the moral distress the agent experiences — and which she judges to be appropriate — does pick out wrongdoing. In his essay "Moral Luck," Williams articulates an emotion closely related to guilt which he calls "agent-regret," and which he characterises as follows:

> [T]here is a particularly important species of regret, which I shall call 'agent-regret', which a person can feel only towards his own past actions (or, at most, actions in which he regards himself as a participant). In this case, the supposed possible difference [between agent-regret and regular regret] is that one might have acted otherwise, and the focus of the regret is on that possibility, the thought being formed in part by first-personal conceptions of how one might have acted otherwise… . It can extend far beyond what one intentionally did to almost anything for which one was causally responsible in virtue of something one intentionally did.
> (Williams 1981, pp. 27-28)

Agent-regret, as Williams defines it, is an emotion which is appropriate when it picks out situations in which an agent has acted in a manner which she in some way finds regrettable. It differs from regular regret in its intentionality; as the phrase "agent-regret" suggests, it is directed towards the

agent herself and some action she has performed. Williams identifies the emotion of agent-regret in the context of an analysis of resultant moral luck; the moral distress one feels in a case of resultant moral luck is plausibly not guilt but agent-regret. But it seems equally possible that it could be agent-regret that the agent experiences in cases of apparent UMW. If this is the case, then Walzer's and Williams's arguments do not meet the burden of proof required to dismiss (or drastically modify) the intuition RCW.

1.4.7: Remaining Questions

Williams's and Walzer's accounts are valuable in that they identify the remainder that exists in cases of apparent UMW. The phenomenology of this remainder involves both an affective component and a cognitive component. The affective component is one of moral distress, and the cognitive component is an intuitive belief that the moral distress is justified. This intuitive belief in the appropriateness of the emotion is held by both the agent himself and by third party observers. For this reason, accounts that dismiss the phenomenological argument for UMW on the basis that the moral distress is unjustified run into problems. However, Williams's and Walzer's arguments do not succeed in establishing that the appropriate distress one experiences in situations of apparent UMW is dispositive of wrongdoing. As Gowans puts it,

> defenses of the phenomenological argument [such as those of Williams and Walzer] have suffered from a tendency to follow a more descriptive and less explanatory approach. Specifically, they have usually been content merely to describe situations in which persons inescapably feel some form of moral distress, and they have been less inclined to go on to give a general normative account that explains why these feelings make sense.
>
> (1996, p. 206)

In other words, Williams and Walzer describe the phenomenology of situations of apparent UMW, but do not give an explanation or argument for its indicating wrongdoing. Rather, they let the phenomenology "speak for itself." (Gowans 1996, p. 205) But this is problematic as Williams's own later work points to another possible explanation for the moral distress — that it is agent-regret. Williams and Walzer show us that the phenomenology of situations of apparent UMW *could* pick out wrongdoing, but they do not show that this is the best explanation. In the next two chapters, I look at arguments to buttress the claim that the better explanation is that it picks out wrongdoing.

Conclusion

In this chapter, I have outlined the reasons which make plausible the two conflicting intuitions RCW and MDJ. I first argued that RCW was plausible. Due to its abstract status it is unlikely to be the result of one of the four forms of bias I identified; and it coheres with another very plausible abstract intuition, the Control Principle. I then argued that MDJ was also plausible. I used the arguments of Bernard Williams and Michael Walzer to illustrate its plausibility, but held that their arguments were not decisive in settling the question, and that if we were to abandon RCW and conclude that UMW was possible, additional arguments would be necessary. In the following two chapters, I look at two accounts which provide these additional arguments.

Chapter Two: Gowans's "Responsibilities to Persons" Account

Introduction

In the previous chapter, I established the starting point of this investigation: the two conflicting intuitions about the possibility of unavoidable moral wrongdoing. I also clarified my understanding of the term "wrongdoing," and explained why I held that the two intuitions were *prima facie* plausible. In this chapter, I aim to show why one ostensibly compelling account reconciling the two intuitions ultimately fails: that of Christopher Gowans (1994). In his sophisticated book length argument for the view that unavoidable moral wrongdoing (UMW) is possible, Gowans contends that while it is not the case that there are true dilemmas — that is, it is not possible that an agent ought to do both *a* and *b* if it is not possible for her to do both — a proper understanding of how moral obligations arise shows us how it is nevertheless possible for an agent to do wrong despite performing the all-things-considered morally right action. The basis of his argument is the view of moral obligations he terms the Responsibilities to Persons (RTP) view.

I proceed as follows. In section one, I sketch out Gowans's argument in detail. Then, in section two, I show why his argument fails to go through. I focus on four problems with the RTP view, and conclude that these problems fatally undermine Gowans's arguments for the possibility of unavoidable moral wrongdoing.

2.1: Responsibilities to Persons: Gowans's Argument

In this section, I lay out Gowans's argument for the possibility of UMW. I begin by summarising Gowans's normative approach, which he names "reflective intuitionism." Then, I lay out his argument against what he calls the Dilemmas Thesis (DT.) This is necessary because it plays a role in his solution to the problem of UMW. After that, I sketch out Gowans's argument that the Responsibilities to Persons account (RTP) provides the best explanation of the phenomenology of situations of apparent unavoidable moral wrongdoing. Accepting this account means accepting that UMW is possible.

2.1.1: Gowans's Reflective Intuitionism

Gowans begins from the same position as I do: from the premise "that our moral intuitions … are likely to have something to do with the truth." (1994, p. 39) That is, he takes our moral intuitions as the starting point for ethical inquiry. He warns that we cannot simply take these intuitions as picking out moral truths; rather, they are "prima facie credible but fallible." (1994, p. 19) In order to ensure that they pick out truth, they need to be brought into a state of coherence or equilibrium. For this reason, Gowans dubs his normative approach "reflective intuitionism," in acknowledgement of what he sees as his debt to John Rawls.[50] But unlike Rawls's method of reflective equilibrium, Gowans's approach does not begin with any foundational principle other than that our intuitions have *prima facie* credibility:

> [T]he use I want to make of this method [of reflective equilibrium] does not imply a commitment to the model of the original position or to the priority of the right over the good. My interest only concerns the idea of bringing into coherence intuitively plausible considerations.
>
> (Gowans 1994, p. 33)

All principles, or "intuitively plausible considerations", are up for revision; there are none that are kept out of the process of reflection and testing.

Gowans notes that we have both abstract moral intuitions (such as OiC) and moral intuitions that emerge through moral experience (or lower-level intuitions). In keeping with his method, he holds that we are unwarranted in seeing either of these types of intuitions as taking precedence over the other, and so when they conflict the correct response is not to summarily dismiss the lower-level intuition, but rather to try to bring the two intuitions into coherence, both with each other and with our other relevant moral intuitions:

> … [I]t is a mistake to give methodological priority either to moral experiences or to abstract principles. The correct approach is to suppose that our intuitions concerning both moral experiences and abstract principles have some initial credibility, and to strive to develop a coherent account of these intuitions—an account that may require modification or rejection of some of them and that includes an explanation of intuitions pertaining to particular cases.
>
> (Gowans 1994, p. 94)

Gowans then moves on to relate this method to the problem of unavoidable moral wrongdoing. He notes we have two conflicting intuitions about the possibility of UMW, one more abstract in nature,

[50] See Rawls (1971, esp. pp. 40-46.)

and one that emerges through moral experience. Because they conflict, they cannot both be correct. One or both will have to be discarded or modified to bring them into coherence. Therefore, we need to test these intuitions not only against each other, but against our array of other relevant moral intuitions and inferential moral beliefs. The best explanation is the one that most plausibly accommodates this array of intuitive and inferential beliefs.

2.1.2: The Options Thesis: Gowans's Explanation for RCW

Having laid out Gowans's normative method and how it applies to the problem of apparent of UMW, I now outline the first step of his argument for the possibility of UMW: that there can be no true moral dilemmas. This is important, Gowans holds, for it explains the intuition that one cannot do wrong in doing what is all-things-considered morally right.

True moral dilemmas are situations in which "the correct conclusion of moral deliberation includes both OA and OB,"[51] and these obligations conflict. (Gowans 1994, p. 49) Gowans dubs this view the Dilemma Thesis (DT). He argues that DT is false and that we should instead prefer the Options Thesis (OT), which is as follows:

> **OT:** For every moral conflict, the correct conclusion of moral deliberation includes exactly one of the following: (1) OA, (2) OB, or (3) ~OA & ~OB & O(A v B).
>
> (Gowans 1994, p. 49)

In other words, in situations of conflict, either one obligation overrides the other, or there is a disjunctive requirement to do either A or B (with the decision as to which to do being made based on nonmoral reasons or else arbitrarily.)

Gowans's argument for this rests on a form of prescriptivism. He holds that the nature of morality is that it is practical and not merely theoretical. If it were not practical, then the distress one felt at the action not taken in a situation of moral conflict would make no sense. This is because if it were merely theoretical it is hard to see why one would care whether one acted on one's moral requirements. Therefore, Gowans argues, morality must be concerned at least in part with generating action-guiding judgements, or judgements that prescribe a particular course of action.

---

[51] Here, OA and OB mean "an obligation to do A" and "an obligation to do B" respectively, where A and B are actions.

Gowans then considers the consequences of this. He rules out what he calls "action-prescriptivism" — the view that to form a moral judgement is to act in accordance with that judgement — because he believes that akrasia (weakness of will) is possible, and therefore that it is possible for an agent to act in contravention of an all-things-considered moral judgement. Instead, he posits what he calls "intention-prescriptivism," which states that "[i]f an agent S sincerely believes that S morally ought to do an action X, then S intends to do X." (Gowans 1994, p. 69) This makes it irrational for an agent to believe that he ought to do both *x* and *y* when he is unable to do both. That it is irrational follows from two principles which Gowans believes flow from "the teleology of intending." (Gowans 1994, p. 70) These are as follows:

> PRI1: If an agent S knows S cannot do an action X, then it would be irrational for S to intend to do X.
>
> PRI2: If an agent S intends to do an action X and intends to do an action Y, then it would be irrational for S not to intend to do both X and Y.
>
> (Gowans 1994, pp. 70-71)

If true dilemmas were possible, an agent would have to intend to do X and intend to do Y, despite the fact that she could not do both. Because of PRI1, this means the agent would have to intend to do both X and Y. But because the agent cannot do both, it would be irrational for an agent to intend to do both X and Y. As OT does not require any irrationality, it is the better explanation of situations of moral conflict.

On this basis, Gowans holds that there can be no true dilemmas; there is always a conclusion to moral deliberation that does not involve conflict. It is this that explains the intuition RCW–it is true that the right conclusion of moral deliberation cannot also be wrong. However, Gowans argues that there is another sense of wrongdoing, one which allows for the possibility of unavoidable moral wrongdoing. It is to this second sense of wrongdoing that I turn in the next section.

2.1.3: The Phenomenological Argument

In the previous section, I laid out Gowans's argument for OT. In this section, I outline his position on why, despite this, unavoidable moral wrongdoing is possible. In short, despite the fact that the correct conclusion of moral deliberation cannot be wrong, there is another sense of moral wrong that allows for UMW.

Gowans begins from the position that the phenomenology of situations of apparent UMW should not be discarded due to its apparent conflict with an abstract principle. This is in keeping with his method of reflective intuitionism (see 2.1.1 above), which grants both abstract and lower-level intuitions *prima facie* credibility. Therefore, in situations of moral conflict, the intuition that one's moral distress at the action not taken is justified deserves to be taken seriously. One might argue that its *prima facie* credibility should be withdrawn due to the fact that it conflicts with the all-things-considered conclusions of moral deliberation. However, Gowans holds that "it is merely an assumption that there is something in the logic of the concept of moral value which imposes this condition. It is neither self-evidently true nor supported by common usage." (1994, p. 90) To dismiss the intuition on this basis, then, is unwarranted. It is entirely possible that, even when one makes the all-things-considered right moral choice, there is a moral remainder that is indicative of wrongdoing. Gowans calls this the Remainders Thesis (RT) and states it as follows:

> OT is true, but there are moral conflicts in which, whatever the agent does, he or she will do something which is morally wrong in the sense of transgressing some moral value.
>
> (1994, p. 91)

If RT is true, then it is the case that one can perform the all-things considered morally right action, and yet do wrong. In other words, the intuition RCW is misguided.

Gowans notes that RT certainly coheres with our moral experience. However, he holds that one cannot simply move from this to the conclusion that RT is true. To do this is to fall into the trap that Williams and others have fallen into — of "rel[ying] too heavily on the description of moral experience… [when] more attention needs to be paid to providing an adequate explanation of this experience." (Gowans 1996, p. 200) That is, an argument is required to show that the phenomenology of situations of apparent UMW picks out actual wrongdoing. Gowans provides such an argument. It can be laid out formally as follows:

P1. There are situations where one feels moral distress even if one does what is all-things-considered morally right.
P2. We have an intuition that this distress is appropriate.
P3. The explanation of this intuition that best coheres with our other intuitions is that it picks out wrongdoing.
P4. For any intuition, we should adopt the explanation that best coheres with other intuitions.
C1. Therefore, moral wrongdoing can be unavoidable.

C1 follows from the four premises, and so if they are correct his argument is sound. P1 and P2 are justified through our moral experience. P4 is justified through Gowans's normative method of reflective intuitionism, a method I agree with. This leaves P3.

Gowans argues that P3 is justified through a normative moral theory which he dubs the Responsibilities to Persons (RTP) account. In the next section, I outline this account.

2.1.4: The Responsibilities to Persons Account

The Responsibilities to Persons account (RTP) plays a key role in justifying Gowans's conclusion that unavoidable moral wrongdoing is possible.

The RTP account begins from the premise that every human being is unique and therefore has unique value. This is evident, Gowans holds, from the fact that when someone dies, we consider it to be an irreplaceable loss. When Antigone dies, for instance, Haimon is distraught, and Creon's suggestion that he can easily find another woman to take her place is rightly ridiculed. This contrasts with an object such as a television set, which, if it stops working, we can fully replace through purchasing another. There is nothing the old television set offered that is not replicated by the new television set. This contrast in our attitudes can only be understood, Gowans holds, "by supposing that each person has unique value." (1994, p. 125)

The RTP account also holds that we have a set of moral responsibilities to those with whom we are in some form of relationship, and that we know these responsibilities intuitively, with the intuition emerging "through the specific, concrete experience of encountering" particular people in particular circumstances. (1994, p. 128) Note that here, Gowans uses the word "responsibilities" to mean "obligations," rather than in the backward-looking sense of meaning something like, "what one

is answerable for."[52] The nature of these responsibilities is contingent on the type of relationship and the unique qualities of each person. As we form relationships with other people, we develop intuitions about "what is involved in [these] particular, concrete relationships with persons with whom we are to a greater or lesser extent, and in various ways, intimate, especially relations of kinship, friendship, and love." (Gowans 1994, p. 122)

Although these intuitions teach us that we have responsibilities to those close to us that we do not have to people with whom we are not intimate, this does not mean that those close to us have greater value than strangers. Using a (gender-reversed) example from Williams, Gowans argues that when a woman chooses to save her husband over a complete stranger, it is not that her husband has greater value than the stranger. It is that she has different responsibilities towards her husband than towards the stranger, in virtue of the relationship that she has with him. The intrinsic value of the two is equal, but her moral obligations towards the two are not. (1994, p. 126)

We do not have any kind of relationship with the overwhelming majority of people, but this does not mean we have no responsibilities towards them. On the basis of the intuitions about our moral responsibilities to those close to us, we can infer moral responsibilities to those who are strangers to us:

> Responsibilities situated in intimate relations form an exemplar on the basis of which moral responsibilities in other circumstances ought to be understood via analogical reasoning from comparisons and contrasts.
>
> (Gowans 1994, p. 122)

We cannot know the unique value of people who are strangers to us, but we know that they *have* unique value (because everyone has unique value); and the fact that they have unique value makes them "the kind of being for whom one can have moral responsibilities." (1994, p. 127) Gowans does not specify exactly what responsibilities we have towards strangers, but he seems to believe that we are able to infer them through consideration of the class or group to which the stranger belongs:

> … a child suffering from malnutrition is deserving of food, simply by being an intrinsically and uniquely valuable creature with this vital need, and quite independent of whatever factors give rise to, or inhibit the alleviation of, the lack of nutrition.
>
> (1994, p. 127)

[52] Elsewhere in this thesis — particularly in Chapters Four-Six — I use the term "responsibility" in this second sense. In this chapter, however, I follow Gowans in using "responsibility" interchangeably with "obligation."

Here, Gowans argues that it is the fact that the stranger belongs to the class of "child" that allows us to see that she is morally deserving of food, and that provides the basis for establishing moral responsibility on the part of other agents. But simply inferring that such a responsibility exists does not settle the question of who actually has the responsibility. Gowans holds that "which persons have the responsibility [in such a case] … depends on a variety of factors," (1994, p. 127) although he does not elaborate on what these factors are.[53]

Finally, Gowans holds that our responsibilities to persons are comparable, but not convertible (1994, pp. 148). That is, he contends that it is possible to compare our moral obligations, and therefore, when they conflict, to determine which is overriding for the purposes of action-guidance. If this was not the case, he notes, then OT could not be true. However, he draws a distinction between comparability and convertibility. In other words, while we are able to compare conflicting moral obligations in order to determine which one ought to guide action, it is not the case that the moral values from which the obligations derive are able to be converted into some common currency or reduced to some more fundamental moral value.

2.1.5: Why Unavoidable Moral Wrongdoing Is Possible

We are now able to see why Gowans believes that it is possible to do wrong despite making the right all-things-considered decision. It has to do with the nature of moral responsibilities on the RTP account.

On the RTP account our moral responsibilities do not emerge from higher level principles (or some "Covering Law," as Gowans puts it) that include "an *a priori* constraint against conflicting responsibilities," but rather "arise from our perceptions of the value of the separate and distinct persons with whom we are connected." (Gowans 1994, p. 134) This means that in certain circumstances our responsibilities to various people will conflict. In such cases, there will be an all-things-considered morally right course of action; but because the conflicting obligations are inconvertible, even if the agent does what is all-things-considered morally right there will nevertheless be a remainder. Furthermore, according to Gowans, this remainder indicates wrongdoing.

---

[53] Martha Nussbaum makes a similar point: "... [I]f people have entitlements, then there is a duty to secure them, even if it is difficult to say to whom the duty belongs. I argue that the whole world is under a collective obligation to secure the capabilities to all world citizens, even if there is no worldwide political Organization. How to assign the duties to specific groups and individuals is a difficult matter… ." (2011, p. 167)

Gowans therefore holds that he has shown how the intuition RCW can be reconciled with the position that the intuition MDJ picks out wrongdoing. RCW is true in the sense that the right conclusion of moral deliberation cannot also be the wrong conclusion. However, there is a second aspect to the concept of wrongdoing which is distinct from the aspect that picks out a failure to act in accordance with the correct result of moral deliberation. When an agent has conflicting responsibilities to persons, the agent does wrong in not acting upon one of the responsibilities even when he makes the right choice all-things-considered, and so MDJ picks out wrongdoing. Gowans therefore provides a solution to the apparent conflict between these two intuitions which does not require the abandonment of either of them.

In this section, I have outlined Gowans's argument for the possibility of unavoidable moral wrongdoing. In the next section, I show why it does not go through.

2.2: Problems with Gowans's Account

In the previous section of this chapter, I gave an outline of Gowans's account of how UMW is possible. His account has several features with which I agree. These include the following:

1. That intuitions–both concrete and abstract–are a source of moral knowledge;
2. That arriving at moral knowledge requires bringing intuitions into a state of equilibrium;
3. That the Dilemma Thesis (DT) is false; that is, it cannot be the case that one ought to do both A and B when it is not possible to do both A and B;
4. That there is (probably) no algorithm or set of rules for determining how to resolve a situation of moral conflict; and
5. We are reliant on intuition in order to resolve such conflicts.

Despite these points of agreement, however, Gowans's view has significant problems. In this section, I aim to show that these problems render it unpalatable as a solution to the problem of unavoidable moral wrongdoing.

Above (see 2.1.3), I stated Gowans's argument as follows:

P1. There are situations where one feels moral distress even if one does what is all-things-considered morally right.

P2. We have an intuition that this distress is appropriate (elsewhere, I have termed this intuition MDJ).
P3. The explanation of this intuition that best coheres with our other intuitions is that the distress picks out wrongdoing.
P4. For any intuition, we should adopt the explanation that best coheres with other intuitions.
C1. Therefore, moral wrongdoing can be unavoidable.

The argument is valid, and so to dispute it one must object to one or more of the premises. Because I accept P1, P2, and P4, I will in this section take issue with P3. [54]

In order to show that P3 is false, one must show that there are insufficient reasons to accept that the best explanation of the intuition MDJ is that it picks out wrongdoing. As Gowans's argument for the idea that the intuition MDJ picks out wrongdoing rests on the RTP account, if one wishes to dispute Gowans's solution to the problem of unavoidable wrongdoing, one needs to object to the RTP view.

In the sections that follow, I make the case that Gowans's RTP account of moral requirements is flawed. I put forth four main arguments for this view. First, I explain that the RTP account entails an implausible relationship between obligations and "oughts." Second, I argue that the account conflicts with the normative approach to which Gowans commits himself, reflective intuitionism. Third, I contend that the inconvertibility of value, which Gowans admits is critical to his argument that the RTP view allows for unavoidable moral wrongdoing, is highly questionable. Finally, I argue that the RTP view cannot make sense of situations of apparent unavoidable moral wrongdoing that do not involve obligations to persons. Because the RTP view grounds Gowans's claim that the best explanation of our intuitions in cases of apparent UMW is that the agent has done wrong, if RTP is unconvincing then we lack grounds to believe that unavoidable moral wrongdoing is possible. I therefore hold that Gowans's argument fails to establish that unavoidable moral wrongdoing is possible.

2.2.1: Obligation and Ought

One reason to view Gowans's view with scepticism is that his conception of moral responsibilities is an odd one. Gowans does not use "responsibility" in the backward-looking sense of picking out some

[54] Note that I would adjust P4 slightly to read, "For any intuition, we should adopt the explanation that best coheres with our other intuitively and inferentially derived beliefs."

outcome or result for which an agent is in some sense answerable.[55] Nor does he use it (only) in the general sense in which one is, for instance, responsible for one's child, or for the company of which one is the CEO. Rather, he seems to use it to mean "obligation" or "moral duty":

> The responsibilities to persons account maintains that there are situations in which a person S has a responsibility to do A and a responsibility to do B even though S cannot do both A and B (hereafter a "conflict of responsibilities.")
>
> (1994, p. 139)

As can be seen, Gowans here refers to responsibilities to perform specific actions; the word "responsibility" could be exchanged with "obligation" without changing the meaning of the sentence. Therefore, it seems reasonable to hold that when Gowans talks about "responsibilities to persons" he means "obligations to persons." But if this is correct then it appears that, if one has a responsibility to a person, then one ought to act in accordance with that responsibility. This is because, plausibly, if one has an obligation to do X, then one ought to do X.[56] But Gowans does not think this; on Gowans's view, one can have a responsibility to a person and yet ought not to act in accordance with this responsibility. Such a situation would arise if an agent S had two responsibilities, X and Y, but could not fulfil them both, and the correct conclusion of moral deliberation was for S to act in accordance with responsibility X. Gowans's view, then, requires us to take the strange position that an agent can have a moral obligation, and yet ought not to act on that obligation.

Of course, if one were to adjust this view to state that an obligation that does not generate a deliberative "ought" statement was a mere *prima facie* obligation, then the view would become unproblematic; such a view is common amongst deontologists, particularly those with an intuitionist metaethics.[57] However, this is not Gowans's position. As I explained above, while Gowans argues that in cases of conflicting responsibilities there is an all-things-considered right result of moral deliberation (and so there are no true moral dilemmas), the responsibility not acted upon is not merely *prima facie*. Rather, it exists as a remainder, and this remainder exerts moral force. Its existence is, on Gowans's view, indicative of wrongdoing.

---

[55] It is in this backward-looking sense that I use the word "responsibility" later in this thesis.

[56] Singer famously notes as much: "'I have an obligation' means no more, and no less, than, 'I ought to.'" (1972, p. 233) Schroeder offers a view which appears more consilient with Gowans's position, holding that "you can have an obligation to do something, even though it is not the case that you ought to do it (because you have a weightier contrary obligation)." (2010, p. 10) This seems to accord with Gowans's view. However, by "obligation" here Schroeder appears to mean "prima facie" obligation. But on Gowans's view, there can be obligations which are not merely prima facie, but which one ought not to act upon. This seems to set his view apart.

[57] See, for instance, Huemer (2005.)

The fact that Gowans's view offers an unusual perspective on the relationship between obligation and deliberative "ought" statements does not mean that it is incorrect. However, it does give us reasons to approach his view with scepticism, and to demand very convincing evidence before accepting it.

2.2.2: "Responsibilities to Persons" as Inconsistent with Reflective Intuitionism

One of the problems with the RTP view is that it appears to conflict with the normative method to which Gowans commits himself, reflective intuitionism. As outlined above, reflective intuitionism requires that lower-level intuitions be brought into coherence with higher-level (or abstract) intuitions, with intuitions at each level being accorded equal importance. Recall that Gowans writes the following:

> [I]t is a mistake to give methodological priority either to moral experiences or to abstract principles. The correct approach is to suppose that our intuitions concerning both moral experiences and abstract principles have some initial credibility, and to strive to develop a coherent account of these intuitions… .
>
> (1994, p. 46)

However, on the RTP view, much greater priority seems to be given to lower-level intuitions. On the RTP account, our moral obligations are the result of a particular person's unique value plus the nature of the relationship which an agent has with this person, and we become aware of these responsibilities intuitively, through the experience of the relationship. It is concrete, lower-level intuitions, then, that are primary in determining our moral responsibilities. This is emphasised by the fact that Gowans believes that moral obligations other than those we have to intimates are known inferentially:

> It is mainly in the context of such relationships that we first come to employ and understand moral considerations. Responsibilities situated in intimate relations form an exemplar on the basis of which moral responsibilities in other circumstances ought to be understood via analogical reasoning from comparisons and contrasts.
>
> (1994, p. 122)

Gowans seems to attribute primary importance to concrete intuitions, with abstract principles merely derivative from these intuitions. This conflicts with reflective intuitionism in a couple of related ways. First, the fact that concrete intuitions are seen as paradigmatic or primary fails to afford equal status to abstract intuitions. Second, even when Gowans acknowledges that abstract moral

principles play a role in arriving at moral knowledge, he appears to see these principles as inferentially derived. But, as I noted earlier (see 1.1), the accepted view of intuition is that it is a non-inferential source of knowledge; an intuition is an initial intellectual appearance. Therefore, not only do abstract principles play an insufficient role in Gowans's RTP account (given his normative commitments), but it also seems that these principles are not intuitive as the term is typically understood in moral philosophy.

A further problem is Gowans's argument that the RTP account shows that there must be a remainder in cases where obligations conflict is unjustified. As discussed above, Gowans contends that, because moral obligations are not derived entirely from an abstract principle (or "covering law"), there is no reason to think that in cases of conflict one of the obligations must be entirely overridden. But it is not clear why he thinks this. After all, despite the fact that — given his own professed normative approach — his view gives too little weight to abstract moral principles, it does not eschew them entirely:

> [Concrete intuitions that emerge in regard to those with whom we have some relationship] should not be understood as a source of self-evident or infallible truths, nor as independent of influence by "theory." Thus my experience of the value of a new acquaintance may be affected by my conception of the value of persons in general. There is clearly a place for general considerations in moral thought. We interpret particular experiences in terms of these considerations. Conversely, general considerations are themselves interpreted and revised on the basis of particular experiences. Such reciprocity is what the method of reflective intuitionism would lead us to expect.
>
> (1994, p. 129)

In other words, lower-level intuitions are shaped or constrained by higher-level intuitions (and vice versa). But if this is so, then it seems as if the abstract intuition RCW ought to constrain our "responsibilities to persons" such that UMW is impossible. This is a far simpler way of bringing our intuitions into equilibrium than the method Gowans actually pursues, which involves positing that we have hitherto failed to appreciate that the concept of moral wrongdoing involves a component that is separate from the all-things-considered action-guiding one.

Some comments Gowans makes later in the chapter could provide the basis for a response to this objection. He contends that "[moral] responsibilities are ways of being that cannot be reduced to a collection of required acts." (1994, p. 133) In addition to a moral responsibility picking out a requirement to perform (or refrain from performing) some action, it also plays a role in determining "the kind of person we are, [and says something] about the value of the person with whom we are related." (1994, p. 133) Because a change in the nature of a moral responsibility towards a person often changes the nature of our relationship with that person, we have reason to "resist the option of

simply abandoning or abridging" a responsibility in a case in which moral responsibilities conflict. (1994, p. 133) Gowans claims that to resist in this way is an "affirmation of the particular value of [the relevant] person and of our standing with them." (1994, p. 133) In other words, then, if a moral responsibility that one was unable to act on became a mere prima facie responsibility, the nature of the relationship with whomever we had the responsibility towards would be weakened. This gives us reason to see responsibilities as persisting, even when it is impossible for us to act upon them, and so provides support for Gowans's view that the intuition RCW should be accommodated through positing a previously unrecognised sense of "wrongdoing."

If Gowans, in talking about one agent's "responsibility" to another, meant responsibility in the general sense I mentioned earlier (that is, in the sense in which one is responsible for one's child), then his arguments here about responsibility changing in response to circumstances would be convincing. But because he seems to use the term to mean obligation, his view is problematic. Gowans is clearly correct that if in cases of conflicting obligations we merely shrugged off the overridden obligations, this would weaken our relationships with those to whom we had the overridden obligations. Take, for instance, the scenario Gowans sketches out involving Craig and the two young women. (1994, p. 98) In this example, Craig and his friend Roberto are hiking in the woods but become separated. While searching for Roberto, Craig slips and breaks his arm. Stumbling on in the hope of coming across other hikers, Craig presently comes across a dirt road and picnic area. He decides to wait there, hoping that other people will materialise who can offer assistance. Eventually, two young women show up in a car, and get out clutching some bottles of beer. Craig approaches them to ask for assistance, but his untidy appearance and air of desperation scares the young women, and they rush back towards their car, obviously intent on leaving. Worried that the young women do not appreciate the seriousness of his and Roberto's plight (Craig reasons that it could be days before other people show up, and Roberto is not an experienced outdoorsman), and that they will simply drive off and not alert the authorities, Craig rushes towards the women and grabs the keys from one of them, so that he can take their car and get help. In this situation, Craig has two moral obligations. One is to get help for Roberto (and arguably to get medical attention for his own injury); and the second is to refrain from taking the young women's car without their consent. However, plausibly the former obligation overrides the latter; while the young women will be aggrieved and inconvenienced if Craig takes their car, this harm does not seem as morally significant as the harm that could befall Roberto and Craig himself if he does not take their car. If this is correct, then Craig's actual moral obligation in this scenario is to take the young women's car. But imagine that, in explaining his conduct later on, someone asked him about his obligations to the young women, and

he said, "Oh, that was only a prima facie obligation, and so I don't need to worry about that." If he were to respond in such a manner, we would find this highly objectionable, and — were they to hear his remarks — the young women would likely respond with outrage. This offers support for Gowans's arguments on the function of moral responsibilities in that it suggests that, intuitively, we do not consider obligations that are non-action-guiding to be eliminated.

But this intuitive reaction can be explained in another way. It seems plausible that what our intuition in response to Craig's hypothetical comment actually represents is our moral censure of Craig for not recognising that his (non-blameworthy) conduct in taking the young women's car generated subsequent obligations. It is uncontroversial that in cases of conflicting obligations, even if the agent performs the all-things-considered morally right action, he may have to take subsequent actions in response to the harm he has (non-culpably) caused. But this does not mean that the non-action guiding obligation in a situation of conflict is not overridden; and so there is no justification for the conclusion that the agent has done wrong.[58][59]

2.2.3: The Inconvertibility of Value

Even if we accept the Responsibilities to Persons view, however, it is unclear that it gives us reasons to think that unavoidable moral wrongdoing is possible. Because Gowans holds that the "inconvertibility [of moral obligations] … is necessary for inescapable moral wrongdoing," (1994, p. 145) if it can be shown that there are reasons to reject the idea that moral obligations are inconvertible, there is good reason to reject Gowans's conclusion that accepting the RTP view commits one to accepting that unavoidable moral wrongdoing is possible. In what follows, I contend that Gowans's arguments for the inconvertibility of moral value beg the question.

On Gowans's view, inconvertibility is when, given a choice between two things of value, "there is something that the poorer choice would have provided that is not provided by the better choice." (1994, p. 148) Gowans gives the following example to illustrate the concept of inconvertibility:

> Suppose I have a choice between two investments. One gives a return of $50,000 and the other gives a return of $100,000. In all other respects (security, liquidity, etc.) they are the same. The second investment is obviously the better choice. Not only is it the better choice, it is a

[58] Later in this thesis (see Chapters Four and Five), I elaborate on this line of argument, building on Wolf's concept of the "nameless virtue" in order to show that the intuition MDJ is best understood as picking out an obligation to repair harm for which we are causally (but not morally) responsible.

[59] It might be objected here that, even if Craig recognized his obligations to the young women, we would still object if this recognition were not accompanied by an emotion of distress. I address this concern later in the thesis (see 6.2).

> choice that gives me everything the first investment would ($50,000) and more (another $50,000). Now suppose my investments have paid off and I am redecorating my apartment. There is a space above the couch that needs to be filled, and two paintings at Christie's have caught my eye, a Degas and a Kandinsky. Either would be perfect for the room, but in entirely different ways. (The cost looks to be about the same, and I do not regard paintings as investments, so the only considerations are aesthetic.) Here too there may well be a better choice, the Kandinsky let's say. Yet, unlike the first example, the Kandinsky does not give me everything the Degas would, and more. Each painting creates an altogether different room.
>
> (1994, p. 148)

Gowans argues that, on his view, moral obligations more closely resemble the paintings than the two amounts of money. Even if an agent, faced with a choice between acting on moral obligation A and acting on moral obligation B, decides that acting in accordance with A is all-things-considered morally right, she nevertheless misses out on something. A does not give her everything that B would have and more. The fact that in such a situation the agent inevitably misses out on something, even when she has performed the morally correct action, gives us (according to Gowans) reason to think that UMW is possible.

Gowans argues that while moral obligations are inconvertible, this does not require that the moral values from which the obligations derive are inconvertible:

> Still, it is important to see that on my account it is not the fact that conflicting responsibilities involve distinct kinds of value that makes wrongdoing inescapable. Rather, it is the fact that they involve distinct responsibilities (which may or may not be of the same kind)... .
>
> (1994, p. 146)

> [it is not the case that] inconvertibility requires pluralism, and hence that inescapable moral wrongdoing requires moral pluralism. Inconvertibility does require a plurality of something, but it need not be a plurality of types.
>
> (1994, p. 149)

But, given the analogy he uses (and which I have quoted above) an inconsistency arises. In this analogy, the inconvertibility lies in the fact that the Degas and the Kandinsky have different types of *value*. This is what justifies his claim that the home decorator loses out on something when she chooses the Kandinsky over the Degas. Thus, his analogy shows that, on a value pluralist view, values are inconvertible (although, as I shall later argue, it is doubtful it even succeeds at this task.) But if as Gowans suggests his view is that conflicts in moral value can arise even on a monistic view, then his analogy is irrelevant to whether moral value is inconvertible.

Gowans anticipates an objection like this. He writes the following:

> But pluralism is not necessary for inescapable moral wrongdoing as I understand it. It is possible for there to be a monist view in which there are conflicting responsibilities… [T]he principle to aid one's friends when in need could generate the responsibility to help this friend and the responsibility to help that friend when it is not possible to do both. Wrongdoing might then be inescapable, but the conflicting responsibilities would be based on the same moral principle. Moral pluralism is thus not required for inescapable moral wrongdoing in the sense of violating a moral responsibility no matter what.
>
> (1994, p. 146)

In other words, Gowans argues that it is possible for a single moral value (such as aiding one's friends) to give rise to inconvertible obligations. But on this view it is unclear why the obligations are inconvertible. If a single moral value is giving rise to the obligations, then the obligations ought to be convertible. A more plausible understanding of the (supposed) fact that an obligation to aid friend A and an obligation to aid friend B are inconvertible is that the situation is one in which two moral values are at stake. That is, aiding A is one moral value, and aiding B is another. It is due to the fact that the moral values conflict that the inconvertibility arises.

In fact, Gowans appears to concede this point when he states the following:

> There is no value of honesty over and above our particular responsibilities to specific persons to be honest. The responsibilities to persons account is Aristotelian in this sense. Wrongdoing, inescapable or otherwise, always involves a violation of a moral responsibility to some particular person, social entity, or what have you. Moral pluralism can be given an Aristotelian interpretation, and so understood I have no objection to it.
>
> (1994, p. 146)

But if this is Gowans's view of moral value, it is hard to see how the view is a monistic one. If "there is no value of honesty" in an abstract sense, but only insofar as it relates to particular people, then his view simply collapses into pluralism.

In order for Gowans's view to work, then, it must be understood as a (moral) value pluralist view. But perhaps if we take his RTP view in this way, his argument about inconvertibility can provide some support for the view. This argument is similar to the Rational Regret Argument in the literature on value pluralism.[60] The argument holds that even when we make a choice we think is all-things-considered right, we experience a sense of distress at what we have missed out on. But if the values each of the choices offered were commensurable — or convertible, to use Gowans's terminology — it would be irrational to experience this regret. Because we have the intuition that such regret is not irrational, the values must be incommensurable (or inconvertible.) But the purpose to which Gowans is putting this argument is problematic. This is because the intuition is very similar to the intuition I

[60] See Hurka (1996), Moen (2016) and Blum (2023.) Stocker (1990) also puts forth a version of this argument in order to support his view that "dirty hands" situations are possible.

have dubbed MDJ, whose conflict with RCW comprises the problem of unavoidable moral wrongdoing. Recall that MDJ and RCW are as follows:

> **MDJ:** In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision.
>
> **RCW:** An action that is all-things-considered morally right cannot also be morally wrong.

Indeed, Gowans is offering up the intuition as parallelling MDJ. But note that he is using this intuition in order to support his RTP account; and the purpose of the RTP account is to justify the view that UMW is possible. Therefore, Gowans is begging the question — he is using the very intuition whose legitimacy we are attempting to reach a conclusion on in order to support a normative theory whose purpose is to justify accepting the intuition. If I am correct, then this is a serious problem for Gowans's argument, as it appears to imperil the view that moral values are inconvertible, and therefore to remove a key pillar of support for the RTP account.

There is a second problem with Gowans's argument from inconvertibility. Gowans asks us to accept that, while values can be "inconvertible", they can nevertheless be compared. That is, it is possible to establish which value is greater. But this is highly dubious. To be able to compare *x* and *y* is to hold that a "better than" relation exists between *x* and *y*. Plausibly, in order for one to decide that *x* is better than *y*, one must first convert *x* and *y* into some common currency; as Crowder puts it, there must be some sort of "super-value" to which the apparently different values *x* and *y* can be reduced (1994, p. 295). But if such a conversion is possible, then it is unclear how *x* and *y* can be inconvertible.[61] This gives us some reason to think that the distinction Gowans makes between comparability and convertibility is unsupportable. As according to Gowans the inconvertibility of moral values on the RTP account is critical to its showing how UMW is possible, this gives us reason to doubt his view.

Gowans might object that the Rational Regret Argument shows that (at least some) values are both inconvertible and comparable. This argument rests on the supposed fact that an agent, when

[61] Some axiologists reject this view. Elizabeth Anderson (1993), for instance, holds that the notion of hierarchical incommensurability can explain how we are able to compare incommensurable values. She contends that, despite the fact that there is no common scale according to which we can compare values, they can nevertheless be arranged in a hierarchy. It is this hierarchy that allows us to resolve cases of conflict between values. But if values are not reducible to some "super-value", it is unclear how we can arrange them in a hierarchy. If such a hierarchy is possible, it seems as if values are commensurable after all. See Williams 2011, section 2 for a more detailed discussion.

choosing between $100,000 and $50,000, could not rationally regret choosing the $100,000, but that he could rationally regret choosing a Kandinsky painting over a Degas (even though he thought the Kandinsky was best.) But this could be countered as follows: In getting only the $100,000, the investor *does* miss out on something–the $50,000.[62] It is correct that an agent could not rationally regret choosing the $100,000 over the $50,000; but it is not correct to hold that the agent could not rationally experience some negative emotion at the fact that he does not get the $50,000 *as well as* the $100,000.

Gowans could respond that although in both cases the choice of one option rules out the possibility of having the other, the difference is that the $100,000 contains the $50,000, whereas the Kandinsky does not contain the Degas. Therefore, even though the choice of the $100,000 leads one to miss out on something (the $50,000), an exact copy of the thing one misses out on is included in the thing one gets. This is not the case when one chooses the Kandinsky over the Degas. There is therefore a reason for one to regret choosing one painting over the other that one does not have when one chooses $100,000 over $50,000.

But this is, once again, begging the question. To hold that the value of the Degas is not included in the value of the Kandinsky is to assume that the value of the Degas (and of the Kandinsky) is inconvertible. As this is the conclusion that the argument aims to establish, this is to beg the question.

Gowans could counter that in the case of the home decorator choosing between the paintings, the emotion of regret or loss would be stronger. This, he could claim, supports the view that paintings have inconvertible value. But, even if we accept that the emotion of regret in the case of the home decorator would be stronger, there is another more likely explanation for this fact. Plausibly, the fact that the emotion of distress is stronger in the home decorator is a result of the vagueness in the value of the two paintings, and a resultant lack of certainty on the part of the home decorator as to whether she has made the correct choice, all-things-considered. To see this, we can apply a principle stated by Ole Martin Moen (2016): the Nominal-Notable Commensuration Principle (NNC). Moen states it as follows:

> [W]hen we seek to commensurate two kinds of values, then no matter how different they are, we always seem to be able to arrive at a clear verdict when one of the values in question is made sufficiently large while the competing value is made sufficiently small.
>
> (2016, p. 1380)

[62] Moen makes a similar point: " if we are forced to choose between giving subject *A* 10 units of pleasure and giving subject *B* 8 units of pleasure, then the better option (giving *A* 10 units of pleasure) does in fact lack something of value that the lesser option has: *B*'s 8 units of pleasure." (2016, p. 1378)

In order to illustrate this principle, let us alter Gowans's home decorator scenario somewhat so that the agent is choosing not between two artistic masterpieces (that is, two notables), but between a Kandinsky and a work by an emerging local artist (a notable and a nominal). In this scenario, the agent judges that the work by the emerging local artist has some aesthetic merit, and she appreciates its intricate brushwork and use of colour. Nevertheless, she without hesitation chooses the Kandinsky; and subsequently feels only a small sense of loss, similar to that of the agent who has to choose between two investments. But if Gowans's view were correct, the agent would experience a much greater sense of loss than would the investor. This is because even though the work by the local artist would have less value than the Kandinsky, its value would nevertheless be inconvertible; and so despite having made the clear best choice, the agent would still miss out on a type of value that she would have had, if she had chosen the work by the local artist. [63]

Therefore, I conclude that his arguments about the inconvertibility of value are unconvincing and do not provide us with reasons to accept the RTP view. As the RTP view is the explanation Gowans offers to justify P3, we lack grounds to accept P3, and therefore also lack grounds to accept that unavoidable moral wrongdoing is possible.

2.2.4: The Incompleteness Problem with RTP

Another weakness of the RTP view is that it fails to provide a complete explanation of the situations in which the intuition MDJ arises. Let us call this the Incompleteness Problem. For instance, RTP provides no way of explaining the intuition MDJ when this intuition emerges in situations that do not involve people. For instance, imagine a situation in which I go hiking in the wilderness and somehow become stranded there, without means of communicating with the outside world (perhaps like Christoper McCandless, for instance, I venture into the Alaskan wilderness during low season, in search of solitude, and run into difficulties.[64]) I have good reason to think that another hiking party will come across me within a month; but I know I will not survive until then given my lack of access to food. Fortunately, I have brought my two dogs with me on the hike; and if I kill one of them and eat it, this will give me enough food to survive until another hiking party turns up. (Let us further stipulate that one or both of my dogs are likely to die anyway from starvation.) Plausibly, in this scenario, the all-things-considered morally right decision is to kill one of my dogs and consume it, to

[63] For another argument along these lines, see Broome (2000.)
[64] McCandless's story is detailed in Jon Krakauer's book *Into the Wild* (1996.)

allow me to survive.[65] However, after I do this, it is highly likely the phenomenology of my experience will include immense moral distress, as well as the intuition that this distress is justified. But Gowans's RTP account cannot provide any explanation of why this is. Gowans's view focuses only on responsibilities to *persons* and how these are (or can be) incommensurate, and so create the possibility of unavoidable moral wrongdoing. But in the scenario I sketched out above, it is responsibilities to animals that are generating the intuition MDJ. This suggests that Gowans's view has an Incompleteness Problem; it does not provide a full explanation of the phenomenology of situations of apparent unavoidable moral wrongdoing.

This problem is one Gowans briefly acknowledges. He writes that although it is in the context of relationships with intimates

> that we come to best understand what is involved in the notion of [moral] responsibility… we clearly have responsibilities in other contexts and of other kinds as well… [These include] responsibilities to strangers within our community, to persons in foreign countries, to human beings on the edges of life such as fetuses or those in an irreversible coma, to generations past and future, to animals, to the environment, to a deity, and so on.
>
> (1994, p. 129)

Gowans holds that it is "beyond the scope of this discussion to consider all such possibilities." (1994, p. 129) Therefore, Gowans could agree that the scenario I sketched out above involves unavoidable moral wrongdoing, but that it is a situation of unavoidable moral wrongdoing that falls outside the scope of his theory. But if this is the case, then it is unclear why we should take P3 (as articulated in 2.2) to be correct. Recall that P3 states the following:

P3. The explanation of this intuition [that is, the RTP account] that best coheres with our other intuitions is that the distress picks out wrongdoing.

If Gowans's view cannot provide a full explanation of the intuition–if it can only explain the intuition that emerges in cases involving people–then it is unclear why we should take his explanation to be the *best* explanation of the intuition. And if we do not have reason to take it to be the best explanation of the intuition, then Gowans's argument falters. Of course, Gowans might argue that his view could be fleshed out or adjusted so as to ameliorate the Incompleteness Problem. However, the onus is on Gowans to show that this is so. In the absence of such an adjustment, the Incompleteness Problem, in

[65] Even a staunch animal rights activist such as Peter Singer acknowledges that, if the gain to human beings is great enough, it overrides the obligations we have towards animals (1975.)

combination with the other issues I have raised above, gives us reason to reject P3, and therefore to reject Gowans's solution to the problem of unavoidable moral wrongdoing.

Conclusion

In this chapter, I have laid out Gowans's arguments in favour of the thesis that unavoidable moral wrongdoing is possible. Gowans dismisses the possibility of genuine moral dilemmas, arguing that it cannot be the case that an agent ought to do A and B despite not being able to do both. There is always, Gowans argues, an all-things-considered right result of moral deliberation. Therefore, it is not the case that wrongdoing can be unavoidable in the sense that one ought to do A and B, despite not being able to do both. However, Gowans contends that there is another aspect to the concept of wrongdoing which has previously been unrecognised: the violation of "responsibilities to persons." Because these responsibilities arise from inconvertible moral values, Gowans holds that when we violate these responsibilities, we do wrong in some sense. However, as I have shown above, the RTP view has significant problems and these problems give us reason to question the idea that it is the best explanation of the intuition MDJ. P3 from my formal restatement of Gowans's argument is as follows:

> The explanation of this intuition (MDJ) that best coheres with our other intuitions is that the distress picks out wrongdoing.

Because we do not have good reasons to accept the RTP view, Gowans's justification for this premise fails. Therefore, we are left without good grounds to accept that unavoidable moral wrongdoing is possible.

Chapter Three: Tessman's "Sacred Judgements" Account

Introduction

In the previous chapter, I analysed Christoper Gowans's compelling argument for the possibility of unavoidable moral wrongdoing. Although it put forth a sophisticated account of how the intuitions RCW[66] and MDJ[67] can be reconciled, serious scrutiny revealed problems that were sufficient to warrant dismissing it. In this chapter, I turn to a more recent attempt to argue in favour of the possibility of unavoidable moral wrongdoing: Lisa Tessman's "sacred judgements" account (2015, 2017). As Tessman explicitly states that her account is in part an attempt to build on what she sees as weaknesses in Gowans's view, it is possible that her refinements allow for the reconciliation of RCW and MDJ in such a way as to sustain the view that unavoidable moral wrongdoing is possible.

Tessman begins from the same starting point as Gowans and I do: by acknowledging the importance of the fact that "[a] moral dilemma [is] significant, even — or especially — when the encounter involves emotions that point in a different direction than the action-guiding deliberations used to choose which moral requirement to satisfy." (2015, p. 12)[68] That is, she recognises that the problem of unavoidable moral wrongdoing results from a clash between the abstract intuition that, if one performs the right action all-things-considered, one cannot do wrong (RCW); and the concrete intuition that arises in certain situations that even when one has done the right thing all-things-considered, one nevertheless experiences justified moral distress (MDJ).

Also like Gowans, she departs from Williams (1981), who takes the intuition MDJ as *conclusive* of the fact that "an overridden moral requirement has survived rather than been eliminated," (2017, p. 30) and that UMW is therefore possible. To sustain the position that UMW is possible, she argues, one must "explain why a moral requirement would (or would not) survive even when overridden; that is… [one must] say what about a moral requirement makes it function" in this way. (2017, pp. 30-31) One cannot simply accept the intuition as authoritative.

---

[66] RCW: An action that is all-things-considered morally right cannot also be morally wrong.

[67] MDJ: In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision.

[68] Tessman uses "moral dilemma" to refer not to what is often called a "true dilemma" — a situation in which neither option outweighs the other–but rather to refer to situations in which, even though one performs the all things considered morally best action, there is nevertheless a moral remainder. (2015, p. 15-16)

In order to show that the phenomenology of situations of apparent UMW picks out actual wrongdoing, Tessman employs a two-part strategy. First, she relies on findings in experimental moral psychology to explain how situations can arise in which it appears that one can do wrong no matter what one does. Then, she focuses on the normative question of why we ought to accept that, in such cases, people are guilty of what she calls "inescapable moral failure."

The chapter proceeds as follows. First, I lay out Tessman's argument for the possibility of unavoidable moral wrongdoing. Subsequently, I analyse her argument. I demonstrate that, despite its ostensible plausibility, the view has serious problems which justify discarding it.

3.1: Tessman's Account

3.1.1: The Dual Process View of Moral Judgements

In this section, I detail Tessman's account of how it is that agents can have the phenomenological experience characteristic of situations of apparent unavoidable moral wrongdoing. On Tessman's view, it has to do with the fact that we have two distinct systems for producing moral judgements.

Tessman relies on research in moral psychology. Specifically, she draws on research by Cushman, Young, and Greene (2010) and Haidt (2001) on the dual process model of moral judgement. According to this model, moral judgements are produced by two processes, one of which is a "reasoning process" and the other an "intuitive," affect-laden process. Moral judgements made according to the former are like "currency," in that the relevant moral values may be weighed against each other and traded off in cases when these values conflict. When "everything relevant has been considered and factored in," a final authoritative or all-things-considered moral judgement has been reached. (Tessman 2015, p. 74) When an agent uses this process to produce such a judgement, there is no moral remainder. This is because the values one is weighing up are commensurable, like currency; in deciding to act on one set of moral reasons, there is nothing that the agent misses out on by not acting on the other set.

In contrast, intuitive moral judgements derive from what Cushman, Young, and Greene call "alarm-bell emotions," which "are designed to circumvent reasoning, providing absolute [moral] demands and constraints on behaviour." (Cushman, Greene and Young 2010, cited in Tessman 2015, p. 70) These moral judgements are not based on any "justifying reasons." (2015, p. 74) Rather, they stem from automatic emotional responses, and built into these responses is a sense of absoluteness.

As Tessman puts it, such responses "issue non-negotiable commands." (2015, p. 71) As an example, Tessman cites a scenario from Greene et al. (2004) called *Crying Baby*, involving a group of people who are hiding from enemy soldiers. The group includes an infant, and the infant begins to cry. This will almost certainly give away the group's hiding spot, and the only way to prevent this is to quickly smother the baby. The "powerful negative affective response"--the intuitive process response–is not only that smothering the infant is wrong, but that it is non-negotiably so. (2015, p. 67) Indeed, taking such an action is "unthinkable" (2015, p. 120) — merely critically reflecting upon the judgement that smothering the infant is right seems to involve a moral transgression. Engaging in such critical reflection would be to have, as Willliams puts it, "one thought too many." (1981, p. 18)[69] Tessman calls such non-negotiable values "sacred values." (2015, p. 94)

According to Tessman, moral judgements arrived at through the reasoning process are based upon inferences and are propositional. In contrast, the intuitive process is arational; that is, it is not reliant on inference. Furthermore, Tessman claims that moral judgements that derive from the intuitive process are like what Tamar Gendler has dubbed "aliefs," which consist of an "associative repertoire of representational, affective, and behavioural components." (Gendler 2010, cited in Tessman 2015, p. 78). According to Tessman, an alief is not a propositional attitude. For instance, if an agent is confronted with a piece of fudge shaped like faeces,

> …instead of saying that I alieve that 'the feces-shaped fudge would be disgusting to eat,' I must say something that will convey all of the representational, affective, and behavioural components of the "associative repertoire" of an alief; so, for example, when presented with the feces-shaped fudge, I could say that I have an alief with content something like: 'dog shit, disgusting, refuse to eat.'
>
> (2015, p. 76)

It is not entirely clear whether Tessman considers the alief to *be* the moral judgement that arises from the intuitive process, or whether she believes that we *derive* the moral judgement from the alief. If the former, then it is hard to see how an alief can be non-propositional, since all judgements are surely propositional. As Gendler herself is not committed to the view that aliefs are necessarily non-propositional, however, one could amend Tessman's view so as to allow that aliefs are propositional, thus ironing out this apparent wrinkle.[70] However, a more charitable interpretation of her view is to

---

[69] This echoes G.E.M. Anscombe's view that it cannot be an "open question" whether executing an innocent person for some greater benefit is morally permissible. For Anscombe, even considering the question exhibits a "corrupt mind." (1958, pp. 16-17)

[70] For instance, in discussing the nature of an alief, Gendler characterises it in part as "the representation of some object or concept or situation or circumstance, perhaps propositionally, perhaps nonpropositionally… ." (2008, p. 643)

adopt the latter reading: that moral judgements are derived from aliefs. There is some evidence for this reading in her work; for instance, Tessman writes that "[o]ne reaches the moral judgement, in each of these aliefs, not through inference, but rather through an associative link." (2015, p. 82) A plausible reading of this is that the alief is the basis of the moral judgement, in much the same way as, on the intuitionist view I endorsed in Chapter One, intuition gives rise to belief. Furthermore, in summarising the view of Cushman, Young, and Greene (2010) in support of her argument, she writes that the emotional response "gives rise to … intuitive moral judgements… ." (2015, p. 70) Therefore, it is plausible that on Tessman's view the intuitive process is arational in the sense that the moral judgements it produces are not the products of inference, but rather arise from automatic, emotional responses which are similar or identical to Gendler's aliefs.[71]

It is because some moral judgements are arrived at through this intuitive process, Tessman argues, that one can experience the sense that unavoidable moral wrongdoing is possible. One way in which moral conflict can arise is when an agent "respond[s] both automatically and through controlled reasoning [that is, through both the intuitive and reasoning processes], with the two processes generating opposite judgments." For instance, in the case of *Crying Baby,* an agent could generate both an "automatic intuitive response ('don't hurt the baby!') and the reasoned response ('the baby will die either way … and saving some lives is better than saving no lives… .')" (2015, p. 84) But this is not the only way in which conflict can arise; it is also possible for "the experience of conflict between two moral judgments … [to] arise from intuitive processing alone." (2015, p. 86) This is because two "alarm-bells" might sound simultaneously, each conveying a different moral judgement that is experienced as non-negotiable (for example, the situation Sophie faces in *Sophie's Choice*.) In such a situation, it will appear to the agent that whatever she does, she violates some moral requirement.

It is now clear why Tessman thinks it is possible for an agent to have the *experience* that, whatever she does, she does wrong. However, as she recognises, the fact that it is possible to have the experience does not mean that we ought to accept the experience as authoritative. In the following section, I lay out Tessman's normative argument.

[71] This view is reinforced through what Tessman has written elsewhere. For instance, in her later book on the same topic Tessman asks us to "[s]uppose there's a man who makes a first-order moral judgment based on an alief such as 'two men kissing each other in public/ disgusting!/ express righteous indignation.'" (2017, p. 105) Notably, Tessman here explicitly states that the judgement is "based on" the alief, not identical to it.

3.1.2: Sacred Judgements and Unavoidable Moral Wrongdoing

In this section, I sketch out Tessman's argument for why we ought to accept the view that an agent can be in a situation of unavoidable moral wrongdoing. Tessman aims to demonstrate that not only is it possible for agents to experience the sense that they face impossible moral requirements, but also that they ought to take this experience as authoritative.

Tessman's case for this view begins with the metaethical constructivist account of moral normativity provided by Sharon Street (2008, 2010, 2012). Metaethical constructivism is understood as the view that moral values are not mind-independent, but rather depend for their existence upon our being the kinds of creatures that engage in acts of valuing. As Street puts it, "things are valuable because we value them, and not the other way round." (2012, p. 41) But this does not mean that reason plays no role in the determination of moral value, for there is a distinction between valuing and desiring. If one desires *x,* and *x* entails *y*, it is nevertheless perfectly coherent for one not to desire *y.* However, if one *values x*, and *x* entails *y*, then it is incoherent not to value *y*. Therefore, moral value is constructed from the "raw materials" of initial normative experience. (Tessman 2015, p. 104) These "raw materials" are then used in a "process of construction." This consists of an agent assessing what Street (2008, 2012) calls the "constitutive entailment" of these initial evaluative judgements — the values that these initial evaluative judgements logically entail — and then attempting to bring the initial judgements and their constitutive entailments into coherence through a process of reflective equilibrium.

Tessman is aware that on Street's view unavoidable moral wrongdoing seems impossible. This is because it requires that, if two evaluative judgements clash, one or both of them need to be modified or eliminated in order to bring about a state of coherence: "[F]or Street, if one violates the requirements of practical reasoning, one cannot be said to be valuing at all." (2015, p. 117) However, Tessman argues that, in requiring that evaluative judgements be brought into coherence, Street imports a normative requirement of rationality that is unwarranted on the metaethical constructivist view. She argues that such an assumption is warranted on a restricted constructivist view,[72] according to which "the substantive judgements that are presupposed are not necessarily thought to be constructed themselves." Therefore, on such a view, "the method of reflective equilibrium … might be the only appropriate way to arrive at evaluative judgements…" (2015, p. 114) However, Tessman argues that there is no reason for a metaethical constructivist to be bound by this constraint. Indeed,

[72] A restricted constructivist view is a view on which not all judgements are subject to the process of construction, with some initial judgements or principles acting as a starting point for the construction process.

she thinks we should reject the assumption, because "there are features of normative experience that indicate that the process of construction does not properly consist in subjecting each evaluative judgment to scrutiny from the point of view of one's other evaluative judgments". (2015, p. 114) On Tessman's view, an equally valid method of ascertaining the authority of evaluative judgements is to take as decisive the quality of non-negotiability that certain evaluative judgements have:

> [A particular evaluative judgment is not] correct by virtue of having eliminated any judgments that conflict with it; [but] rather [because] it has survived by having a quality—based on one's normative experience of being non-negotiably required—that somehow secures its spot as part of one's evaluative standpoint even when it is overridden.
>
> (2015, p. 119)

In other words, there are certain evaluative judgements which carry a quality — that of sacredness — which prevents them from being subjected to the scrutiny of reflective equilibrium (for such scrutiny corrupts their sacredness.) The authority of such judgements therefore rests on the very quality of sacredness which prevents their being exposed to scrutiny. Tessman argues that attributing authority to such judgements on this basis is justified because it better accords with a constructivist view. Street's version of metaethical constructivism is impoverished, Tessman argues, because she "takes the product of construction to be contingent only upon "what one values" while failing to take into account "how one values." (2015, p. 117) Because some of our judgements carry with them the quality of non-negotiability, she holds, we are justified in attributing authority to these judgements. If we are justified in considering some moral judgements to have authority even when they conflict with one or more of our other moral judgements, then we are justified in holding that there can be situations in which, even if one does what one judges to be all-things-considered morally best, one can still violate a moral requirement. Therefore, an agent can do wrong unavoidably.

Tessman is aware that if some evaluative judgements gain their authority automatically, and thus exempt from the scrutiny the method of reflective equilibrium offers, then it appears as if there is no way to reject any such judgement, and all must be deemed authoritative. Tessman acknowledges that, if this were true, it would be highly problematic — it could lead to the acceptance of some sacred intuitions that ought not to be considered sacred. Tessman considers particularly problematic the possibility of sacralising judgements that are "ideologically shaped through social and political exercises of dominating power." (2015, p. 129) However, she claims that there is a way of gaining "some of the protection" that reflective equilibrium offers without ever subjecting the judgements to the process of reflective equilibrium (and thereby violating them.) This involves ensuring that one

should interact with other members of one's moral community who have sacralised (at least some) different values:

> When other people do not share my intuitive judgments about what is and is not unthinkable, they can take themselves to be free to engage in reasoned reflection about whether to reject a value that I accept as sacred; they can do the thinking that I have judged it to be impermissible for me to do … [T]heir critical thinking … may serve to reframe an issue for me, or to make it more transparent, and this may reveal to me that I cannot continue to be confident in what I previously took to be a sacred value.
>
> (Tessman 2015, p. 136)

In other words, one ought to interact with people who have not sacralised exactly the same moral judgements as oneself, as they will be able to present arguments against (some of) the values that one has sacralised, or to act in such a way as to show that they do not consider particular values to be sacred. One could not rationally consider whether the reasoning of these people was valid, as to do so would violate the sanctity of one's judgements. However, their arguments and actions may then trigger intuitions in one, and these intuitions may strip some sacred judgements of their aura of sanctity. This, Tessman argues, allows for one's automatic moral judgements to change, and offers some of the protection that reflective equilibrium provides one with normally.

Finally, Tessman does not contend that it is always wrong to violate a judgement that appears sacred. Rather, she uses Gowans's view and ideas from Martha Nussbaum to lay out the circumstances in which clashing moral judgements mean that the agent unavoidably engages in wrongdoing. She argues that Gowans's arguments on the inconvertibility of value (see 2.2.3) "provide a good first clue to how to draw the distinction between moral requirements that can remain standing (and become impossible moral requirements) even when overridden in conflict, and moral requirements that are fully eliminated when rightly overridden…." (2015, p. 38) However, she argues that sometimes there are clashes between inconvertible values that leave no remainder. In order to show which clashes are such, she draws on Nussbaum's notion of "tragic costs."[73] (Nussbaum 2000) According to Tessman, when the all-things-considered morally correct course of action "leaves someone beneath a threshold level of that to which they are entitled," (2015, p. 42) there is a moral

[73] According to Nussbaum, "tragic costs" are those which deny agents the "capability to function" in a list of 10 areas, including life, bodily health, and control over one's environment. When an agent performs an action (or chooses not to perform an action) and which pushes an agent "beneath the threshold on any one of the 10 capabilities," (2000, p. 1023) a tragic cost has been incurred and the agent inflicting the cost does wrong. Tessman does not commit herself to Nussbaum's list of (as Nussbaum puts it) "Central Human Capabilities," but rather uses Nussbaum's view to give plausibility to the idea that costs above a certain threshold leave a remainder indicative of wrongdoing, even when the person who inflicts the costs takes the all things considered morally correct course of action.

remainder and the agent unavoidably does wrong. But there is no such remainder when the action does not cause this threshold to be breached. To illustrate this, Tessman imagines that she had discarded her daughter's favourite toy because "it had been recalled for safety reasons." (2015, p. 39) Such an action would be morally correct, but would nevertheless cause her daughter to lose something she held to be uniquely valuable. However, this loss of unique value does not mean that Tessman would be guilty of wrongdoing, because "small enough irreplaceable losses [such as the loss of a favourite toy] should be treated as acceptable." (2015, p. 39) In other words, such a loss would not put her daughter below the threshold of that to which she was morally entitled, and so if the action that led to the loss was all-things-considered morally right, it would not also be morally wrong (or, to use Tessman's terminology, constitute moral failure.)[74]

3.2: Evaluating Tessman's View

In this section, I aim to evaluate Tessman's "sacred judgements" account of why unavoidable moral wrongdoing is possible.

One objection that could be levelled against the account is that metaethical constructivism is false. However, this falls beyond the scope of this thesis.[75] For the purposes of this chapter I set aside any problems constructivism might have as metaethical theory, and assess whether Tessman's view works given that assumption. In what follows, I argue that Tessman's view contains internal contradictions and has highly implausible implications. For these reasons, we are justified in rejecting it. Before this, however, I address an issue relating to the terminology Tessman uses, justifying the view that what she is arguing for is indeed unavoidable moral *wrongdoing*, and that her account is therefore a legitimate candidate for showing that such wrongdoing is possible.

3.2.1: Tessman's Terminology

Although Tessman explicitly acknowledges her debt to Gowans, she opts for a different term to describe the form of moral violation she argues for. Tessman utilises the term "moral failure" instead of "moral wrongdoing" to characterise an agent's breach of an impossible requirement. Therefore, it could be argued that what she has in mind is distinct from the main topic of this thesis, unavoidable

[74] Tessman reinforces this point elsewhere, writing that a "moral requirement can … be canceled if, despite the fact that not fulfilling it will cause unique loss, this loss is the kind of loss that people ought to expect to bear." (2023b, p. 499)

[75] For a recent criticism of Street's brand of metaethical constructivism see Morton (2018.)

moral wrongdoing. Tessman states that she uses "moral failure" because "wrongdoing" might "be taken to suggest something more specific (than the violation of a moral requirement)" such as "the violation of a duty." (2015, p. 14) However, elsewhere Tessman uses "moral wrongdoing" and "moral failure" interchangeably; for instance, she writes that "it is not always wrong" to violate a moral requirement that appears to remain standing even when one is unable to fulfil it. (2015, p. 53)[76] Also, she later states that an agent who has to decide between saving his wife and saving a stranger is inevitably "commit[ting] a wrongdoing." (2015, p. 93) Finally, Tessman uses "wrongdoing" in other work in order to describe the phenomenon. For instance, when discussing soldiers in wartime Tessman writes that an action can be "the best possible thing to do and in that sense was the right thing to do, while in some other sense it was still a wrongdoing." (2023a, p. 35) Therefore, it seems reasonable to suppose that Tessman's notion of moral failure is not very different from what I have termed moral wrongdoing. Indeed, "moral failure" as she uses the phrase appears synonymous with "moral wrongdoing."[77]

3.2.2: Sacred Judgements and Non-Truth-Aptitude

An element of Tessman's account is that (as she acknowledges) it requires that some of our moral judgements be non-truth-apt. According to the metaethics that Tessman (largely) endorses, evaluative truth is arrived at when all one's evaluative judgements withstand scrutiny "from the standpoint of other judgments…" (2015, p. 111), or are in a state of reflective equilibrium. However, because non-negotiable judgements, on Tessman's view, are arrived at through an automatic intuitive process and cannot be subjected to the process of reflective equilibrium, these judgements cannot attain the status of truth, and are therefore not truth-apt. Instead of calling these judgements true, we may instead call them "authoritative."

But this points to a problem with Tessman's account. According to Tessman, a judgement is authoritative when it has been sacralised well, and not authoritative when it has been sacralised badly. She uses the terms "well" and "badly" instead of "correctly" and "wrongly" because using the latter would contradict her view that sacred judgements are not truth-apt:

---

[76] See also later in Chapter One, where Tessman again uses the word "wrongdoing" in this context. (2015, p. 27)

[77] For another example of Tessman using "unavoidable moral wrongdoing" instead of "unavoidable moral failure," see Tessman (2023a.)

> it might sound more natural … to say that sometimes people sacralize values wrongly, but, because I have said that evaluative judgments about what is sacred need not be truth-apt, I will call them "wrong" only by using scare quotes.
>
> (2015, p. 127)

Given that Tessman does not consider sacred judgements to be truth-apt, the question arises of what it could mean to say that a value had been "badly" sacralised. Tessman appears to consider a judgement to have been badly sacralised when the intuition that gives rise to it is the product of unjust social forces:

> [S]ome moral intuitions are ideologically shaped through social and political exercises of dominating power. I take this to be an especially compelling worry, and one that makes it clear how risky it is to exempt some moral judgments from scrutiny through a method such as reflective equilibrium.
>
> (2015, p. 129)[78]

In other words, the intuitive process that gives rise to aliefs on which our sacred judgements are based can be influenced by forces — and therefore advance agendas — that, from the standpoint of our other evaluative judgements, are unjust. There is therefore a conflict between the sacred judgement and an agent's other moral judgements which, if the sacred judgement were to be subjected to the process of reflective equilibrium, would result in its elimination. When such a conflict exists, Tessman's view goes, a judgement can be said to have been "badly" sacralised.

However, if this is what it means to badly sacralise a judgement, then it is unclear what grounds Tessman has for holding that such judgements are not truth-apt. This is because, according to Tessman's metaethics, normative truth is established through the process of reflective equilibrium. If a sacralised judgement has been "badly" sacralised when it conflicts with other evaluative judgements, then it seems as if it is truth-apt, and when it is overridden by other moral judgements we ought simply to call it wrong. If sacred judgements were not truth-apt, then it is unclear how identifying a conflict with judgements could result in their being overridden.

One might respond that Tessman could rescue her view from incoherence by altering it such that sacred judgements *are* truth-apt. But such a move cannot save her view, for Tessman's claim that sacred judgements are not truth-apt is crucial to her contention that metaethical constructivism does not require that they be subjected to reflective equilibrium. It is only because they are not truth-apt that they are exempt from the scrutiny of reflective equilibrium, and so a concession that sacred

[78] For more on the relationship between oppressive ideologies and morality, see Tessman (2005.)

judgements were truth-apt, after all, would amount to a concession that her grounds for holding that unavoidable moral wrongdoing is possible had collapsed.

One final line Tessman could take to defend her view is to question my interpretation of when, on her view, a judgement has been sacralised badly. Perhaps a sacred judgement is not badly sacralised when it conflicts with another judgement, but rather when it is the product of forces which one considers to embody normatively false judgements. For instance, maybe a range of patriarchal moral judgements which an agent considers false influence the automatic intuitive processing system such that it produces intuitions that conflict with other (anti-patriarchal) moral judgements. If this is what Tessman means, then it is not the conflict between moral judgements that allows a judgement to have been badly sacralised, but rather the fact that the intuitive processing system that has produced the sacralised judgement has been improperly influenced.

But it is not clear how Tessman's metaethics can allow for this. On a metaethical intuitionist view (see 1.1), it is indeed the case that forces such as the culture one has been raised in can introduce bias into our intuitions, rendering them untrustworthy, in much the same way as some problem with our visual system might render visual perception untrustworthy.[79] However, because Tessman does not believe that our intuitions are picking out mind-external objective truth, it does not appear as if the factors that influence their production can be of any significance in determining their reliability, except insofar as those factors are judgements that one has reason to reject. But as I outlined above, if a badly sacralised judgement is merely one that conflicts with another moral judgement, then these judgements must be truth-apt. But such a view is incompatible with Tessman's argument for the possibility of UMW, which depends on the claim that some moral judgements are non-truth apt and therefore not subject to being brought into equilibrium with other moral judgements.

3.2.3: Sacred Judgements and the Problem of Authority

A related concern has to do with Tessman's claim that the quality of sacredness can give "authority" to a moral judgement. According to Tessman, Street simply "smuggles in" the assumption that a metaethical constructivist view requires that evaluative judgements withstand the scrutiny of reflective equilibrium. Rather, metaethical constructivism allows for other ways of granting authority to evaluative judgements. Tessman contends that this, coupled with the fact that some of our moral judgements emerge through an automatic, intuitive process, justifies relying on this quality of sacredness to grant authority to certain moral judgements.

---

[79] See Huemer (2008.)

But Tessman's arguments on this point are unconvincing. Tessman seems to believe that we are justified in taking the quality of sacredness as imparting authority to an evaluative judgement because it better recognises how we value:

> [I]t is important that–as empirical work has shown–much of our valuing and evaluatively judging is done without controlled reasoning…. The standard picture of construction as always requiring reasoning fails to recognize that reasoning is not the only (or even the primary) cognitive process through which humans as valuing creatures do what I take to be what construction is really about: the creation of values that one affirms from one's own evaluative standpoint…
>
> (2015, p. 117)

In other words, because some of our evaluative judgements emerge automatically, we are justified in relying upon an automatic mode of conferring authority upon judgements, too. But it is not clear what the relevance is of how we value. Street is in perfect agreement with Tessman about the fact that some of our evaluative judgements are not the product of reason but rather emerge automatically.[80] Consider a thought experiment Street uses to illustrate how normative truth gets started. She imagines that two highly sophisticated creatures of the same species spring (improbably) into being. Due to a genetic mutation,

> they happened to take different things to be valuable. As it so happened, the first valued its own survival and nothing else, whereas the second valued its own destruction and nothing else… When the first creature judged that its own survival was good, and the second creature judged that its own survival was bad, the first was not recognizing some normative truth that the second was somehow missing.
>
> (Street 2008, p. 221)

Note that in this example, the two creatures simply "happen" to have different evaluative judgements. They are not arrived at through any process of reason; it is simply in the nature of each creature to make a particular evaluative judgement. These judgements reflect "mere instances of valuing, born of chance alone… ." (2008, p. 222) In other words, they are arrived at automatically. Because the two creatures do not value anything else, these evaluative judgements cannot come into conflict with anything. But Street notes that, if we "imagine a third creature" who values two things, such as "its own survival and the survival of its offspring," then conflict amongst automatic evaluative judgements is possible, and to determine normative truth the creature would need to bring its judgements into equilibrium. (2008, p. 223) Therefore, evaluative judgements that are the result of

[80] Street (2008, pp. 222-223) calls such judgements about values "non-instrumental," in the sense that they are judgements about what is intrinsically valuable. Judgements about what is instrumentally valuable are arrived at through a reasoning process.

automatic processes are perfectly compatible with the type of metaethical constructivism that Street advocates. In contrast to Tessman's claims, Street's account does not overlook the significance of the fact that some of our evaluative judgements emerge automatically, and so the fact that some judgements derive from an automatic process does not justify the claim that such judgements are off limits to reflective equilibrium.

Indeed, Tessman's own view appears more vulnerable to the charge of "smuggling" normative assumptions into metaethical constructivism. Street rejects the charge that her view "presuppose[s] substantive assumptions about reasons." (2008, p. 231) She holds that, if A values *x*, and *x* entails *y*, then A values *y*, but this does not imply that A *ought* to value *y*. She is arguing that simply *is* (constitutively) valuing *y*.[81] Street offers the following analogy: "Being a parent entails having children, but that does not mean that a 'parent' is making an error if she has no children; she's just not a parent." (2008, p. 232) Street's view therefore does not conceal an illegitimate normative assumption; her ideas on constitutive entailment simply explicate an aspect of what it means to make a normative judgement. But the same cannot be said for Tessman's view. This is because Tessman *does* import an assumption into the process of construction: the assumption that the quality of non-negotiability or sacredness that some judgements carry ought to be taken as authoritative. Therefore, Tessman's criticism seems more apt when applied to her own view than to Street's.

This line of criticism becomes even more potent when we consider Tessman's position on when moral costs should be taken up. As I explained above (see 3.1.2), on Tessman's view, it is only when an agent whose all-things-considered morally right action inflicts a cost of sufficient severity on another agent that there is a moral remainder indicative of wrongdoing. But it is not clear how exactly this requirement comports with the other aspects of Tessman's account. It appears incompatible with the view that an agent could bring the judgement about the severity of costs into reflective equilibrium with her non-negotiable moral judgements by discarding the quality of non-negotiability when acting on the judgement would not produce costs of sufficient severity. This is because, on Tessman's view, it is impermissible to subject non-negotiable judgements to the scrutiny of reflective equilibrium; and modulating them so that they comport with judgements about the severity of costs seems to require exactly this. Therefore, it would have to be a fact about the intuitive process that sacred judgements simply do not emerge except when acting on the judgement would (or would be likely to) impose costs of sufficient severity.

But here the view runs into a problem. It seems as if judgements about what constitute acceptable costs must be derived through reason. Nussbaum's (1990) judgement regarding tragic

[81] This is why Street terms it "constitutive entailment."

costs certainly is; it does not emerge intuitively, but rather through reasoned reflection. Tessman does not commit herself to Nussbaum's view (or to any particular view), but it is hard to imagine that any candidate for replacing Nussbaum's view could emerge intuitively. However, if this is so, then Tessman is importing a foundational assumption into the construction process: the reasoned judgement that certain types of loss or cost is simply "a cost to be borne," (Tessman 2015, p. 42) whereas a more serious loss or cost leaves a moral remainder. This is problematic for Tessman in two ways. First, it means that making non-negotiable judgements is not a brute fact about how we value, but rather a contingent fact; presumably, if we reasoned our way to the judgement that there are no tragic costs, no sacred judgements would emerge from our intuitive process. If this is correct, we have no reason to take the making of sacred judgements to be a necessary element of how we value, and no reason to incorporate it into the method of metaethical constructivism. Second, importing a foundational assumption like this seems to conflict with metaethical constructivism, and makes Tessman's view more like a form of restricted constructivism. Because one of Tessman's arguments for preferring her view over Street's is that her view is a truer version of metaethical constructivism, this is highly problematic.

3.2.4: The Problem of Revising Moral Judgements

A final problem with Tessman's view concerns the way in which it allows for (or fails to allow for) the revision of moral judgements. As I noted above (see 3.1.2), Tessman is aware that her rejection of the view that it is necessary to bring one's judgements into a state of equilibrium appears to make it difficult to alter "badly sacralised" moral judgements. However, she argues that her view allows for another method of revising moral judgements. This method, as I sketched out above, involves deliberately exposing oneself to members of one's moral community who have sacralised different moral judgements.

But this view is problematic, for a couple of reasons. First, the method seems highly unlikely to work in practice. As Regina Rini has noted, it is very improbable that a person, having sacralised a particular moral judgement, would react in the way Tessman supposes upon encountering the representations of those who have not sacralised the value, or indeed oppose it. (2016, pp. 1232-1233) Imagine, for instance, that someone experiences as non-negotiable the moral judgement that interracial relationships are bad. It is quite unlikely that, upon encountering an anti-racism rally, that person would pause to listen attentively to the protestors shouting slogans, and read the banners being waved around, and so on, because she knows that unless she does so she gives herself no way

of modifying badly sacralised moral judgements. Indeed, it seems far more likely that the agent would react in anger to rally and storm off to denigrate the protestors on social media or to others with similar views.

Tessman could respond that while it is true that people *would* likely act in this way, they ought not to. Perhaps it is the case that agents *should* expose themselves to the arguments of those who have sacralised different moral judgements. But this is not a particularly promising line of argument, either. This is because, if an agent considers a particular moral judgement to be sacred, it is unclear how he could endorse a strategy of exposing himself to influences that might result in the rejection of the non-negotiable judgement without thereby violating the sanctity judgement. To illustrate this, imagine that I judge that it would be wrong to cheat on my girlfriend, and the judgement has the quality of sacredness. On Tessman's view, I ought to test this intuition by exposing myself to influences that might challenge this sacred status. But it hardly seems plausible that I ought to try to trigger new intuitions by, say, frequenting bars patronised by numerous attractive young women (therefore exposing myself to sources of potential influence over my sacred judgements regarding my relationship.) Nor does it seem plausible that I ought to ensure that I read tracts justifying cheating on one's partner.

In other words, if a judgement is truly sacred — that is, if one judges that it would be wrong to question or interrogate the judgement directly — it is unclear how one could consider it morally unproblematic to arrange circumstances such that the judgement would be interrogated indirectly. The endorsement of this indirect strategy seems to breach the sacredness of the judgement in the same way as a more direct strategy would. If this argument is correct, then it appears as if Tessman's account fails to provide agents with a morally permissible way of altering moral judgements they consider sacred, and it is unclear how her account could be reconstructed so as to rectify this weakness. This gives us reason to view her account — and with it her claims that unavoidable moral wrongdoing is possible — with a good deal of scepticism.

Even if we set these problems aside, it appears that Tessman's view can only accommodate random change in our morality. Tessman offers up the suggestion above as a way of "correcting" our mistakenly sacralised moral judgements. But it is just as likely that — if this method was able to alter our judgements at all — it would alter them in the opposite direction. Imagine that an agent has sacralised the judgement that abortion is wrong except in cases of rape or incest. However, he is aware that he could have badly sacralised the judgement, and so, in keeping with Tessman's suggestion, exposes himself to the actions and communications of others who have sacralised different values. However, as a result of this exposure, instead of correcting the badly sacralised

judgement about abortion so that he judges that abortion is always permissible, the agent instead revises it such that he holds that abortion is never permissible, even in cases of rape or incest. Such a result seems just as likely on Tessman's view as the kind of correction she would hope for (towards more permissive judgements regarding abortion); and so Tessman's account at best appears only to allow for random moral change, rather than improvement.

Conclusion

Lisa Tessman has offered a detailed and intriguing argument in favour of the possibility of unavoidable moral wrongdoing. Like Gowans's Responsibilities to Persons account (see Chapter Two), it attempts to plug a weakness in phenomenological arguments in favour of unavoidable moral wrongdoing, such as that of Bernard Williams, in that it offers an argument for taking the moral distress that arises in such situations as indicative of wrongdoing. According to Tessman, if metaethical constructivism is true then the view that all normative judgements must be subjected to the process of reflective equilibrium is false. Some judgements derive authority from a quality of sacredness, and it is therefore possible for an agent to be in a situation in which she faces two moral oughts, neither of which is overridden. However, as I have argued in this chapter, the initial plausibility of the view crumbles under detailed scrutiny. First, there are good reasons to reject Tessman's argument that metaethical intuitionism allows for methods other than reflective equilibrium to imbue evaluative judgements with authority. Tessman's contention that some of our evaluative judgements arise automatically provides no clear reason to reject the method of reflective equilibrium; and her argument that bringing one's evaluative judgements into equilibrium is itself a normative principle that may, on a metaethical intuitionist view, be jettisoned, is unconvincing. Second, the view that sacred moral judgements are non-truth-apt conflicts with Tessman's contention that these judgements can be "badly" sacralised. It appears as if a judgement can only be discovered to have been sacralised badly if one subjects it to the scrutiny of reflective equilibrium. But as reflective equilibrium is the method by which (on a constructivist view) normative truth is arrived, this is incompatible with the view that sacred judgements are not truth-apt. Finally, Tessman's view does not provide any clear method for revising sacred moral judgements. Although she argues that exposing oneself to other moral agents who have sacralised different judgements can offer a means of moral revision, there are good reasons to think that deciding to expose oneself in such a way is itself a violation of the sacredness of one's non-negotiable judgements.

In a later paper on apparent unavoidable moral wrongdoing in situations of war, Tessman writes that an account of the phenomenon of apparent unavoidable moral wrongdoing should ensure that "the phenomenon of moral injury should be addressed in a way that both acknowledges taking responsibility as a fitting response and recognizes the value of having shared normative expectations that are fair." (2023a, p. 41) I agree. In the rest of this thesis, I lay out a theory which reconciles the conflicting intuitions that emerge in a situation of apparent unavoidable moral wrongdoing in a way that acknowledges that the agent must take responsibility for the negative consequences of her actions, but denies that she is guilty of wrongdoing.

Chapter Four: The Nameless Virtue

Introduction

In the previous two chapters of this thesis, I analysed two accounts that strove to demonstrate that unavoidable moral wrongdoing was possible. I argued that each of these accounts — one by Christopher Gowans and the other by Lisa Tessman — had serious flaws, and that neither provided a compelling reason to accept the possibility of UMW. This left us where we began: with two compelling intuitions which appear contradictory, and no satisfactory way of reconciling them.

In this chapter, my aim is to provide a solution which allows for such a reconciliation. Recall that the two apparently conflicting intuition are as follows:

> **An action that is all-things-considered morally right cannot also be morally wrong (RCW).**
>
> **In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision (MDJ)**

One explanation of MDJ is that the distress is guilt and picks out wrongdoing. But because this would mean discarding or radically revising RCW, we would need a very compelling argument for the idea the MDJ picks out wrongdoing. As I have argued in the previous two chapters, the most compelling existing accounts do not withstand scrutiny. Therefore, we need to find another way of reconciling the two intuitions.

In this chapter, I argue that these two intuitions can be reconciled when one properly understands the nature of the distress the agent experiences. When it is understood as agent-regret, I contend, one is able to reconcile MDJ with RCW. In making my case, I draw on Susan Wolf's ideas on moral luck (2001), in particular her contention that the problem of moral luck (at least, one type of moral luck, resultant moral luck) can be resolved through a recognition of what she calls the "nameless virtue," and an understanding of the distinction between guilt and agent-regret.

The chapter proceeds as follows. In section 1, I lay out Wolf's account denying resultant moral luck. In section 2, I show how it can be applied to the problem of UMW and argue that this provides us with reasons to doubt whether UMW is possible. In particular, I contend that situations of resultant moral luck and UMW are analogous in that they both produce in the agent a feeling of moral distress,

and that this distress is self-directed and considered appropriate. In both cases, I claim, this moral distress is best explained as agent-regret. That the agent is subsequently required to make amends for the consequences of her actions — despite not being morally responsible for them — is due to what Wolf dubs the "nameless virtue," the virtue of assuming responsibility for what one is merely causally responsible for. [82]

Therefore, the chapter aims to make the case that it is plausible that Wolf's solution to the problem of resultant moral luck also provides a way of reconciling the two conflicting intuitions which constitute the problem of unavoidable moral wrongdoing.

4.1: Wolf's Solution to the Problem of Resultant Moral Luck

In this section, I lay out Susan Wolf's solution to the problem of resultant moral luck (RML.) I do this in preparation for applying it — with some adaptations — to the problem of unavoidable moral wrongdoing.

Moral luck, a term first used by Bernard Williams in a 1976 essay (reprinted in 1981), describes the role that luck seems to play in determining how blameworthy we are for our actions, or the degree to which we do wrong.[83] Nagel (1979, reprinted in 2012), in an essay in response to Williams, identified four types of moral luck: constitutive, circumstantial, resultant, and causal. Constitutive luck concerns luck in the way our characters are constituted. Circumstantial moral luck concerns the way in which the circumstances we find ourselves in seem to influence the way we are morally evaluated. RML concerns how one's moral evaluation depends upon the results of one's actions. Finally, causal moral luck concerns the role of the laws of nature in bringing about one's (morally relevant) actions.

Moral luck is considered a problem because it seems to conflict with the control principle (CP). As expressed by Nagel, this principle states that "people cannot be morally assessed for what is not their fault, or for what is due to factors beyond their control." (Nagel 2012, p. 962) As I noted in

---

[82] In this chapter and what follows, I follow David Miller (2005) in using Hart and Honore's "commonsense" notion of causal responsibility. On this notion, an agent S is causally responsible for event X if S is one of the antecedent conditions necessary to bring about X, and S's contribution is abnormal. For instance, in Williams's famous truck driver scenario, the unusual feature is the truck driver's brakes failing. Cf. Hart and Honore (1985) and Miller (2005, p. 455.)

[83] As has been noted, the typical meaning of the word luck does not seem to overlap entirely with how the word is used in the phrase "moral luck." But this is irrelevant because, as Hartman notes "the moral luck debate is about not luck per se but a tension in our ordinary thinking about moral responsibility." (2017, p. 24) See also Zimmerman (2019, p. 217) and Statman (2019) for more on this topic.

Chapter One (see 1.3.2), this principle seems highly intuitively plausible.[84] This creates a problem as various situations seem to give rise to contrary intuitions, intuitions to the effect that luck *can* play a role in moral evaluations. The murderer whose bullet hits its target seems more blameworthy, and therefore guilty of greater wrongdoing, than the would-be murderer whose bullet misses (a case of RML), or the one whose car breaks down on the way to commit his murder (a case of circumstantial moral luck.)

In her essay, Wolf focuses her arguments only on providing a solution to the problem of RML, or luck as to "how things turn out." (Wolf 2001, p. 5) Wolf holds that the phenomenon of RML appears to be extremely prevalent: "Every day, people in laboratories, government offices, corporations, and universities sign off on things to which they ought to put a stop, or bend the rules for the sake of convenience or laziness or misplaced generosity. Rarely, but occasionally, such acts of flawed reasoning, negligence or weakness blow up in someone's face." (2001, p. 5) These everyday occurrences are examples of RML, Wolf argues, because it is outside the person's control as to whether or not negative consequences result; and so the degree of wrongdoing she commits is also outside of her control. In other words, these situations violate the control principle, "that people cannot be morally assessed for what is not their fault, or for what is due to factors beyond their control." (Nagel 2012, p. 962) Wolf argues that although it appears this way, this is in fact deceptive, and that if one appreciates the existence and role of what she calls "the nameless virtue," the problem dissolves, and it becomes clear that there is no RML.[85]

In order to illustrate the apparent phenomenon of RML in greater detail, Wolf draws on Williams's (1981) example of a truck driver (let us call him Driver_1) who neglects to have his brakes checked and then while driving runs over a young child who darts unexpectedly out onto the street.[86] Had Driver_1 had his brakes checked, the child would still be alive, and so it seems as if he has committed a grave moral wrong.[87] Wolf then asks us to imagine a second truck driver (whom I shall refer to as Driver_2) who also fails to get his brakes checked, but in front of whose truck no child runs out. While we might feel somewhat reproachful towards him for his negligence, we do not experience

---

[84] As Enoch and Marmor (2007) put it, "the condition of control has enough intuitive support and philosophical credentials that one would need very strong reasons to discard it." (p. 407)

[85] The idea that there are certain virtues (and vices) which are nameless can be traced to Aristotle in *Nicomachean Ethics* (See 1108a, where Aristotle argues that an "excess, deficiency and mean" of anger are all "practically nameless." See also 1107b and 1125a)

[86] Note that the truck driver in Williams's original example displayed no negligence at all, and so differs from Driver_1 in important ways. I use WD_1 to refer to Williams's version of the truck driver.

[87] Wolf often (although not exclusively) refers to the agent as "blameworthy" rather than guilty of wrongdoing, as does Nagel (who also uses "culpability."). As I made clear in Chapter One, I take it that to be blameworthy or culpable for something is to have done wrong, and so do not distinguish amongst these usages. For an argument which attempts to decouple blame and wrongdoing, see, for example, Tannenbaum (2018.)

such a strong emotion towards his conduct as we do towards that of $Driver_1$. However, both agents have committed exactly the same moral failing: neglecting to get their brakes checked. It is mere bad luck that a child happened to run onto the street in front of $Driver_1$, for which the driver presumably cannot be held responsible. The problem, then, is how we can reconcile this with our sense that $Driver_1$ is guilty of a greater moral violation than is $Driver_2$.

Wolf notes that some philosophers have adopted what she calls (in a non-pejorative sense) the irrationalist position (that is, a position accepting of resultant moral luck.) This position holds that, despite the fact that $Driver_1$ and $Driver_2$ act identically, $Driver_1$ nevertheless commits a more grievous moral wrong, because his negligence results in a child's death. Others have adopted the rationalist position, holding that despite what our intuitions might suggest, the two truck drivers do equal moral wrong, because they have acted in an identical fashion. Wolf states that the rationalist position "seems to [her], on first reflection, the more justifiable." (2001, p. 6) Nevertheless, it is "positively eerie" (2001, p. 6) to hold that is appropriate for $Driver_1$ and $Driver_2$ to believe themselves to be in the same moral situation. This, Wolf believes, points to a problem with the rationalist position. It is not that it is incorrect, but rather that it is "incomplete". It fails to capture, or even to acknowledge, something morally significant in the phenomenon of moral luck, something that suggests some grain of truth in the irrationalist response." (2001, p. 6) Wolf claims that her account (laid out subsequently) fills this gap in the rationalist account, rendering it complete and unproblematic.

Wolf is aware that a rationalist could claim that the "eerie" quality might be due to epistemic limitations. In a case where an agent's actions cause harm, for instance, we are more likely to become aware of their morally wrong actions and therefore to attribute guilt to the agent.[88] In addition, we tend to judge the morality of an action partly in accordance with its expected consequences, but it is often hard to know the "approximate range of expected harm" (2001, p. 7) of an action. Therefore, we tend to take the actual consequences of the action as indicative of its expected consequences, and so condemn actions with very negative consequences more harshly than those with only mild negative consequences, or no negative consequences at all, even when the actions are identical. If these explanations are correct, then the rationalist account goes through; the "eerieness" Wolf attributes to the rationalist explanation is merely due to our epistemic limitations.

However, Wolf claims that while "imperfect knowledge" might explain in part our different reactions to $Driver_1$ and $Driver_2$, it does not explain the different reactions we expect the drivers to have. In order to illustrate this, she asks us to imagine the reactions of the two drivers in the aftermath of their behaviour and its consequences. The first would feel dreadful about the consequences of his

[88] Richards (1986, pp. 199-201) offers an argument along these lines.

negligence, experience nightmares, and likely find it difficult to resume driving his truck. The second would, upon reading about the first truck driver's accident in the paper, likely hasten to get his brakes checked and resolve to never again be so lax about maintaining his truck. But he would not find it difficult to resume his career as a truck driver, nor would he experience any nightmares or emotional trauma. Wolf claims that these divergent responses on the part of the truck drivers demonstrate the gap in the rationalist account because we consider their respective reactions appropriate. Even if we were to know everything there is to know about both cases, it is not plausible that we would find it appropriate that the truck drivers should react in the same way. If this is correct, then, the rationalist position is an incomplete account of the moral landscape.

It seems, then, that we are back where we started — caught between the control principle and the intuition that the two truck drivers' different responses are appropriate. Wolf's proposed solution to this predicament is to distinguish between moral responsibility and causal responsibility; and also between guilt (which we appropriately experience when we have committed moral wrongdoing) and agent-regret (which is appropriate when we are causally responsible for negative outcomes.) Wolf takes the concept of agent-regret from Williams (1981.) The two truck drivers, Wolf argues, commit an equal moral wrong and are equally deserving of blame. However, the actions of $Driver_1$ — the truck driver who hits a child — have produced terrible consequences for others. It is therefore appropriate that, in addition to guilt, he experience the negative moral emotion of agent-regret, and that he take responsibility for these consequences. This accounts both for our different responses $Driver_1$ and $Driver_2$, and for their own different responses.[89] In other words, it enables us to reconcile the rationalist position with the irrationalist position.

Why is it that in cases of apparent RML one ought to take responsibility for the consequences of one's actions, even beyond what one is morally to blame for? Wolf's position is that this normative expectation is due to the fact that "[t]here is a virtue that … we all dimly recognize and commend that may be expressed as the virtue of taking responsibility for one's actions and their consequences. It is, regrettably, a virtue with no name… ." (2001, p. 13) This virtue, Wolf argues, is "a species of, or at least akin to, the well-established virtue of generosity." (2001, p. 14) Generosity, on Wolf's view, involves "[giving] more—more time, more money, more love, more lenience, more, in one way or another, of oneself than justice requires." As an agent such as $Driver_1$ is not (entirely) morally responsible for the negative consequences that result from his action, to take responsibility is to go beyond what justice demands of him. But, Wolf argues, this is unproblematic, because–as recognition of the nameless

[89] On this view, it is appropriate for $Driver_2$ to experience mild guilt, due to the fact that he has failed in his obligation to properly maintain his truck. Of course, many drivers who fail in this way will not experience any guilt; but this does not mean that it is not fitting for them to experience it.

virtue suggests–it can be the case that we ought to act in ways that go beyond the requirements of justice.

Wolf argues that there is another reason why we expect agents such as $Driver_1$ to act in accordance with the nameless virtue. To illustrate this, she asks us to contrast what we expect of someone who merely witnesses a bad event with what we expect of an agent who is causally responsible for that bad event. We do not expect the witness to offer to pay for the vase. However, we do expect the agent who broke the vase to offer to pay for it. Wolf claims that in both cases, offers of compensation for the vase are examples of generosity, in that they go beyond what justice requires. But the fact that we only expect such an offer from the breaker of the vase shows that the nameless virtue is a virtue that extends beyond justice—it has "another aspect" to it, one Wolf holds is "not so clearly or narrowly moral." (Wolf 2001, p. 14) This other aspect is its salutary effect on what she calls the "psychic health" of the agent. An agent who attempted

> to draw sharp lines between what one is responsible for and what is up to the rest of the world, to try in this way, to extricate oneself and others from the messiness, and the irrational contingencies of the world, would be [an agent who] remove[d] [him]self from the only ground on which it is possible for beings like ourselves to meet.
>
> (Wolf 2001, pp. 14-15)

What Wolf seems to mean by this is that it is not possible for us to live fulfilling lives while disclaiming any type of special connection to what we are causally responsible for. This too, Wolf argues, plays a role in justifying us as seeing the nameless virtue as picking out an actual virtue.

Wolf is aware that, if her account is correct, it would seem to leave a part of the initial problem intact. A key part of the irrationalist approach is that, if two agents act identically but one causes worse consequences than the other, the agent who causes worse consequences will experience greater guilt than the other. Wolf's account would seem to provide no way of accounting for this, however, other than to say that he is experiencing an inappropriate level of guilt or that he is mistaken about his emotions (that is, he is unable to distinguish between guilt — an emotion of moral distress whose cognitive component is something along the lines of, "I have committed moral wrongdoing" — and agent-regret, which is also an emotion of moral distress, but one whose cognitive component does not include an endorsement that the agent has committed wrongdoing.[90] Wolf's response to this explanation is that it is partly correct, and that the mistake is understandable. Because guilt and agent-regret are similar emotions, she writes, "when a faulty action causes a harm, our feelings tend to get all mixed together — several sources for feeling bad about oneself and for being motivated to engage

[90] See Chapter Five of this thesis for more on the cognitive component of agent-regret.

in certain sorts of actions combine to produce a mental state that is not itself analyzable into component states." (2001, p. 17) In other words, the guilt and agent-regret merge in such situations, leading the agent to make mistaken estimates about his degree of moral wrongdoing and guilt. This is further understandable in such situations because it is hard to determine precisely how much wrong an agent has done and therefore how much guilt is warranted. In the case of $Driver_1$ and $Driver_2$, for instance, it is unclear whether they should both feel as guilty as the one who ran over the child, or as the one whose negligence produced no negative consequences, or perhaps somewhere in between. Wolf holds that the answer (although presumably determinable) is not immediately apparent, further affecting agents' ability to experience guilt in appropriate quantities and exacerbating the conflation of guilt with agent-regret.

To summarise, then, Wolf's proposed solution to the problem of RML is to hold that agents who act identically commit identical levels of moral wrong. However, an agent who is causally responsible for a negative outcome appropriately experiences a related emotion of moral distress, agent-regret, thus accounting for the greater moral distress in an agent such as $Driver_1$. Furthermore, an agent such as $Driver_1$ is expected to act in accordance with what Wolf calls the "nameless virtue." This virtue grounds a moral requirement: the requirement to take responsibility for the consequences of one's actions even when one is not morally responsible for those consequences (hereafter I shall refer to this moral requirement as RTR). [91]

In the next section, I seek to apply Wolf's ideas to the problem of unavoidable moral wrongdoing, and argue that the nameless virtue can explain why, despite negative feelings towards oneself, one does not do the morally wrong thing in cases of apparent unavoidable moral wrongdoing.

4.2: A Proposed Solution to the Problem of Unavoidable Moral Wrongdoing

In this section, I lay out a case for how Wolf's arguments can be applied to situations of apparent unavoidable moral wrongdoing to resolve the conflict between the intuitions MDJ and RCW. The application of her ideas to such situations, I argue, shows that unavoidable moral wrongdoing is impossible. I contend that the self-directed negative emotion is best explained through Williams's

[91] In a recent and interesting paper, Adam Piovarchy (2020) has made a somewhat similar point about the obligations that accrue to an agent who causes harm blamelessly, although he uses the notion of "excused wrongdoing" to refer to those who are causally responsible for harm: "[W]rongdoers often acquire new duties to try to repair the relationship with their victims, even if they are not deserving of blame for their wrongdoing." (p. 148) He draws on Williams's example of the truck driver as a paradigm case of "excused wrongdoing." But the phrase "excused wrongdoing" seems oxymoronic, and the concept contradictory. If an action is excused, it is unclear how the action can also be wrong.

concept of "agent-regret," and that our intuitions regarding how agents should feel and act in these situations can be explained through Wolf's concept of the nameless virtue.

4.2.1: Similarities Between Situations of Resultant Moral Luck and Situations of Unavoidable Moral Wrongdoing

The situations of resultant moral luck of which Wolf writes have a clear connection to cases of supposed unavoidable moral wrongdoing. I will briefly explicate this connection here.

The first feature that the two situations have in common is that it appears that the agent's degree of wrongdoing is outside of her control. In cases of RML, for instance, A and B act identically, but — for reasons outside of her control — the actions of agent A produce worse consequences than the actions of agent B, and so it appears as if A has committed greater wrongdoing than B. Similarly, in cases of UMW, despite performing the all-things-considered morally right action the agent appears to do wrong. Therefore in both types of situation the way the agent is morally evaluated is (to some extent at least) outside of her control.

Second, the phenomenology of agents in such situations appears similar. In both type of situation the agent experiences deep distress, and this distress is self-directed. Furthermore, this distress seems appropriate, both to the agent herself and to third party observers. In other words, the intuition MDJ arises in both cases.

Finally, it appears that agents in situations of RML and UMW both bear some responsibility for the consequences of their actions. When luck causes a minor case of agential negligence to metastasize into something horrible (as in situations of RML), we expect the agent whose negligence has been transformed by luck to nevertheless assume responsibility for the horrible consequences. Similarly, when an agent performs a morally right action but it results in harm to others (as in situations of UMW), we expect the agent to take responsibility for this harm.

There are good reasons, therefore, to take situations of resultant moral luck and situations of unavoidable moral wrongdoing as similar in the relevant respects. Plausibly, then, the problem of UMW can be solved using Wolf's solution to the problem of RML. In the following section, I show that this is indeed the case — that applying Wolf's solution to the problem of RML to situations of UMW allows us to reconcile the intuition RCW with the intuition MDJ. I'll call this the Nameless Virtue View. (hereafter the NV View.)

4.2.2: The Nameless Virtue View to the Problem of Unavoidable Moral Wrongdoing

In the previous section, I identified key similarities between situations of RML and UMW. Here, I lay out how Wolf's solution to the problem of resultant moral luck points the way to a solution to the problem of unavoidable moral wrongdoing. In order to show that Wolf's ideas can explain situations of unavoidable moral wrongdoing generally (and not merely specific instances), I consider three such situations.

As I noted earlier, the problem of UMW lies in the fact that the intuition RCW clashes with the phenomenology of specific situations of moral conflict. In these situations, the agent experiences an emotion of moral distress which intuitively seems appropriate (an intuition I have termed MDJ). It also seems as if the agent has some obligation to make amends for her behaviour. Advocates for the possibility of UMW argue (for reasons I have laid out in previous chapters) that these factors justify the belief that UMW is possible. As I contended earlier (see Chapters Two and Three), the arguments for this position have problems, and we are therefore not warranted in abandoning the compelling intuition RCW. However, this leaves the tension between the two intuitions unresolved. Wolf's solution to the problem of RML seems to provide a way to resolve this tension. First, let us consider the situation of Sophie, from the novel *Sophie's Choice.* In a concentration camp during World War II, Sophie is accosted by a sadistic Nazi guard, who demands that she must pick which of her children she wants to save (the other will be taken away, presumably to be slaughtered.) Here, she is put in a position in which, no matter what she does, grave harm (in fact, death) will accrue to a person she holds very dear. She makes the all-things-considered morally right decision to sacrifice one of her children (it is all-things-considered morally right because otherwise both would be killed, which is obviously morally worse). Nevertheless, there are horribly negative consequences — Eva, her daughter, is whisked away, and presumably killed.

The NV View allows us to make sense of this situation. We can understand the moral distress Sophie experiences as agent-regret; and see it as appropriate because she is causally (but not morally) responsible for negative consequences.[92] Therefore, we can retain the intuition MDJ. However, Sophie has not committed wrongdoing — her action is all-things-considered morally right — and so applying Wolf's ideas to this situation allows us also to retain the intuition RCW.

[92] One could argue here the Sophie was not causally responsible for what happens to her daughter — it is the guard who takes the child away against Sophie's will. However, this overlooks the fact it is Sophie who chooses which of her children will be abducted by the sadistic guard. It is this act of choosing that generates Sophie's causal responsibility in the situation.

Furthermore, Wolf's ideas allow us to make sense of the intuition that Sophie's decision generates subsequent moral obligations. Such an intuition would appear to clash with the contention that the agent had not committed wrongdoing; but when one recognises the nameless virtue, and the moral requirement it grounds (the requirement to take responsibility), this apparent clash dissolves. When an agent is causally responsible for harm, there is a moral requirement for her to take responsibility for this harm. Admittedly, Sophie is unable to take reparative actions in the same way as someone who is causally responsible for a lesser harm; whatever she does, it will not bring her daughter back. Nevertheless, she can at least try to take actions to ameliorate the wider harm, to family members and to herself.

The NV View allows us, in the case of Sophie, to reconcile the apparently conflicting intuitions RCW and MDJ. That is, it allows us to preserve the intuition that an all-things-considered morally right act cannot also be wrong, while simultaneously allowing us to make sense of the fact that we feel that it is appropriate for Sophie to experience self-directed moral distress. To show that Wolf's ideas apply generally to situations of apparent unavoidable moral wrongdoing, I will now consider her ideas in relation to a couple more such situations.

I will begin with a situation Gowans uses and that I discussed in Chapter Two, involving Craig and the two young women whose car he steals (I have adapted it slightly to make it more realistic in digital age). The situation is as follows: Craig and Roberto are hiking in the woods in an area with no cell phone coverage when they become separated. Craig hunts for Roberto, but cannot find him, and is considering triggering his emergency beacon when he slips and injures himself, and the beacon tumbles away down the side of the mountain. Craig is able to stagger on, and eventually he comes to a picnic area at the end of a dirt road. After a while, a car pulls up at the picnic area, and two young women get out. When he approaches them to appeal for assistance, however, they are frightened by him, and begin hastily retreating to their vehicle. Worried about Roberto and unsure of the severity of his own injuries, Craig grabs the car keys, and drives off to seek assistance. He feels great moral distress at the harm he is causing to the women, but decides that it is all-things-considered the right choice. He drives to the nearest place with cell phone coverage, calls emergency services, and returns to the picnic area, where he apologises to the angry girls and offers to pay for the petrol. He also resolves to buy them gifts to atone for the way his actions have affected them.[93]

Gowans claims that Craig's moral distress is best explained by holding that, despite doing what is all-things-considered best, he nevertheless does wrong. As I argued in Chapter Two, such a view is unconvincing. The NV View appears to provide a more successful explanation. On the NV View,

[93] See Gowans (1994, p. 98) for his synopsis of the hypothetical situation.

the distress Craig experiences is agent-regret; although he has not done anything wrong, he is causally responsible for negative consequences (taking the young women's car and abandoning them in the woods), and his agent-regret is therefore appropriate. The fact that we consider it appropriate for Craig to apologise to the girls and resolve to buy them presents is not because he has done something wrong for which he must atone, but rather because of the requirement to take responsibility–the moral requirement stemming from the nameless virtue to take responsibility for bad consequences that one is merely causally responsible for.

Finally, I will show how the NV View can explain a situation from Chimamanda Ngozi Adichie's novel *Americanah* (2013). In the novel, Ifemelu is a young Nigerian woman who has moved to America to attend university. Because her visa prevents her from working legally, she struggles to find casual employment to fund her studies and cover her living expenses. She is unable even to secure employment in fast food or waitressing. This leads her to accept an offer from a tennis coach to come to his house to "help him relax." She understands that this is a euphemism for sexual activity, and she has no desire to engage in such activity with him; she has a boyfriend back in Nigeria. However, her financial situation is so dire that she is struggling to pay the rent and to feed herself, and so she accepts the tennis coach's offer. After the encounter, she experiences a deep sense of distress:

> Afterwards, she lay still, coiled and deadened. He had not forced her. She had come here on her own. She had lain on his bed, and when he placed her hand between his legs, she had curled and moved her fingers. Now, even after she had washed her hands, holding the crisp, slender hundred-dollar bill he had given her, her fingers still felt sticky; they no longer belonged to her… . Back in her apartment, she washed her hands with water so hot that it scalded her fingers, and a small soft welt flowered on her thumb. She took off all her clothes, and squashed them into a rumpled ball that she threw at a corner, staring at it for a while… Her voice mail light was blinking. It was probably Obinze [Ifemelu's boyfriend]. She could not bear to think of him now.
>
> (Adichie 2013, p. 154)

In this situation, it is plausible that Ifemelu performs the all-things-considered right action. She is already a week late with the rent and is struggling to scrounge together the money to pay for food. Her aunt, also living in America, is unable to provide her with assistance because she is studying medicine, and her "account is almost empty." (Adichie 2013, p. 79) Nor do her family or friends back in Nigeria have the means to assist her. It is therefore a choice between possible homelessness and hunger and a sexual encounter which involves betraying her partner. It is implausible to claim that an agent must allow herself to become homeless and starve on the street rather than accept money for sexual activity and cheat on her boyfriend, and so I will assume that Ifemelu acts morally correctly,

all-things-considered. Nevertheless, her feelings of moral distress (self-directed, as is evident from the fact that she symbolically washes her hands) also seem appropriate.

One could object here that this example is different in a key respect from Wolf's example and the examples of UMW I have previously referred to. This is because it appears that there is no third party who has been harmed, and this suggests that the emotion is not agent-regret (as Ifemelu is not causally responsible for any harm.) Of course, Obinze, Ifemelu's boyfriend, *would* be harmed were he to find out,[94] and so we could describe the moral distress as subjunctive, in that the emotion is triggered by subjunctive morally bad consequences. But this is problematic. To see why, imagine that a driver (call him Lucky Driver) took his improperly maintained truck out and, instead of hitting a child, drove around without incident. It is highly implausible to claim that it is appropriate for him to experience agent-regret because, if he had hit a child, a morally bad consequence would have arisen.[95] Indeed, the idea of subjunctive moral distress appears even more absurd when we consider that any time we do anything some morally bad consequences could arise, and so if the mere possibility of being causally responsible for morally bad consequences were enough to occasion agent-regret, we all ought to experience it constantly. This, one could argue, shows that the NV view cannot the phenomenology of all situations of apparent UMW, and so should therefore be viewed with scepticism.

However, I do not think this objection works. Plausibly, having engaged in sexual activity with another man — even though it was all-things-considered morally right — Ifemelu has an obligation to tell her partner about it; and when she tells Obinze, she knows he will be greatly pained. Therefore, Ifemelu's moral distress is not subjunctive, but rather *in anticipation* of the harm she will cause Obinze.[96] As the one who will be causally responsible for the pain, it is appropriate for her to experience agent-regret even in anticipation of the hurt she will cause Obinze. Anticipatory agent-regret seems equally plausible in cases of RML. Imagine that Driver_1 somehow becomes aware (perhaps he becomes clairvoyant, or God descends to inform him) in advance that he will hit a pedestrian during his journey, but that, for some reason, there is nothing he can do to avoid it. It is likely that Driver_1 will experience agent-regret in the lead up to the catastrophe in anticipation of the harm he will cause. If this is correct, then the NV View can also be applied successfully to explain our

[94] In actual fact, he is harmed, because Ifemelu is so distressed by the encounter that she cuts off all contact with him.

[95] Of course, he ought to experience guilt over knowingly driving around in a poorly maintained truck, but this is the case whether he hits a child or not.

[96] In the novel, Ifemelu does not inform Obinze of her actions. Instead, unable to muster the courage to tell him what she has done, she cuts off all contact with him.

intuition that Ifemelu acts correctly, while also accounting for our intuition that it is appropriate for her to experience moral distress.

Finally, it seems as if Ifemelu is required to make amends for her actions. Given that, *ex hypothesi*, she has acted correctly all-things-considered, this seems puzzling. However, the NV View provides a way of solving this apparent contradiction. Because she is causally responsible for negative consequences, she is required to take responsibility for them (perhaps by apologising to Obinze, explaining the circumstances that led to the sexual activity, and demonstrating to Obinze how much she cares for him.) There is therefore no contradiction between the fact that she has done right, all-things-considered, and the fact that she must take reparatory action.

Conclusion

In this chapter, I have put forth the novel proposal that the contradictory intuitions that arise in situations of unavoidable moral wrongdoing can be reconciled through applying Wolf's solution to the problem of resultant moral luck to such situations. I've called this the Nameless Virtue View (NV View.) The distress an agent experiences is, I have argued, agent-regret; and the way that we expect agents to act in the wake of being causally responsible for bad outcomes is explained by the fact that there is a moral requirement to take responsibility for what one is merely causally responsible for.

However, in order to make this proposal more compelling, numerous questions must be addressed. Primary amongst these are questions concerning the nature of agent-regret and of the nameless virtue. Wolf (understandably, given the brevity of the paper) is somewhat vague on why exactly we should accept that there exists a moral requirement to take responsibility for what we are merely causally responsible for; and there also remain important questions about the nature of agent-regret. In the next chapter, I aim to address these issues.

Chapter Five: Agent-Regret and the Requirement to Take Responsibility

Introduction

In the previous chapter, I argued that Susan Wolf's proposed solution to the problem of resultant moral luck could plausibly be adapted to solve the problem of unavoidable moral wrongdoing, a solution I termed the Nameless Virtue (NV) View. I noted, however, that given the brevity of Wolf's account there were certain gaps in her argument. This is particularly the case in relation to two key components of her solution: agent-regret and the nameless virtue. What reason do we have for taking there to be a moral requirement to take responsibility for what we are only causally responsible for? Why should we accept that the moral distress that agents feel in situations of resultant moral luck (and unavoidable moral wrongdoing) is distinct from guilt? My aim in this chapter is to sketch out answers to these questions, and in so doing plug the aforementioned gaps in Wolf's account.

The plan of the chapter is as follows. First, I examine whether there are plausible grounds for accepting that there is a requirement to take responsibility for what we are merely causally responsible for. Using Wolf's essay as a starting point, I lay out formally the argument for accepting this, and contend that the premises are justified. I argue that there are two forms that this justification might take: intuitive and consequentialist.

Second, I argue that it is plausible that there is an emotion of moral distress distinct from guilt, and respond to an argument from Michael Moore to the effect that there are good reasons to see what is typically classified as agent-regret as guilt, after all. Then I examine the argument that agent-regret as it is typically understood is irrational or inappropriate. As we judge the moral distress in situations of RML and UMW to be appropriate, if agent-regret is irrational then we run into a contradiction. I examine several arguments to defeat this objection that agent-regret is irrational, and find that none is convincing. Then I make my own argument for the rationality of agent-regret, which requires recognising the differences between what Greenspan calls "representational rationality" and strategic rationality. I argue that the rationality of emotional responses to fiction provides us with a model for how, in cases of non-culpable responsibility for negative outcomes, emotions of distress can be rational.

5.1: The Nameless Virtue and the Requirement to Take Responsibility

In the previous chapter I argued that Wolf's ideas could furnish us with a way of reconciling the intuitions RCW and MDJ. In this chapter I provide a defence of some of these ideas. In this section, I offer a defence of the nameless virtue, by focusing on the requirement for action that it grounds — the requirement to take responsibility for what one is causally responsible for (RTR.)

Wolf's justification for accepting RTR is twofold. First, as I have noted, she likens it to generous action. Because generous action is typically seen as supererogatory and not required, this seems to present some problems. However, as I discuss Wolf's claims about RTR's relationship to generosity in the following chapter (see 6.1), I will not discuss this aspect of her argument here. I will instead focus on her second justification for taking the nameless virtue as a virtue: that it is beneficial for one's "psychic health":

> We are beings who are thoroughly in-the-world, in interaction with others whose movements and thoughts we cannot fully control, and whom we affect and are affected by accidentally as well as intentionally, involuntarily, unwittingly, inescapably, as well as voluntarily and deliberately. To form one's attitudes and judgments of oneself and others solely on the basis of their wills and intentions, to draw sharp lines between what one is responsible for and what is up to the rest of the world, to try in this way, to extricate oneself and others from the messiness, and the irrational contingencies of the world, would be to remove oneself from the only ground on which it is possible for beings like ourselves to meet. If we define ourselves in ways that aim to minimize the significance of contingency and luck, we do so at the cost of living less fully in the world, or at least at the cost of engaging less fully with the others who share that world.
>
> (Wolf 2001, pp. 14-15)[97]

Wolf's claim here seems to be that, if one did not acknowledge the "significance of contingency and luck" in shaping how our actions affect others, one would not be able to participate as fully in the activities that make life meaningful (or as Wolf says, it would come "at the cost of living less fully in the world.") However, there are some problems with this justification. First, if it is wrong to "aim to minimize the significance of contingency and luck," then why not simply accept the irrationalist

[97] Wolf is here possibly responding to Nagel, who writes the following: "I believe that in a sense the problem has no solution, because something in the idea of agency is incompatible with actions being events, or people being things. But as the external determinants of what someone has done are gradually exposed, in their effect on consequences, character, and choice itself, it becomes gradually clear that actions are events and people things. Eventually nothing remains which can be ascribed to the responsible self, and we are left with nothing but a portion of the larger sequence of events, which can be deplored or celebrated, but not blamed or praised." (Nagel 2012, p. 975) See also Williams's "Moral Luck," where a similar passage occurs. (1981, p. 29)

position (that there is extant resultant moral luck)? Second, even if we put this matter aside, what does it mean that to fail to accept contingency is to "remove oneself from the only ground on which it is possible for beings like ourselves to meet"? It is true that our actions often have consequences that, given the type of creatures we are, we could not have predicted; but it is less clear why we should accept that these consequences should affect our judgements about others and about ourselves. Why could we not all reject responsibility for that for which we are merely causally responsible, and allow others to do likewise?

I will begin with the first objection: that if we accept a role for contingency and luck in our existence, then it is unclear why we should not simply accept moral luck. In other words, the objection goes, Wolf's argument is simply self-defeating. But we can avoid this objection by recognising that there is a reason for not accepting a role for luck in determining what we are morally blameworthy for: the compelling intuition that "luck in how things turn out is irrelevant to blameworthiness." (Wolf 2001, p. 6) There is no such compelling reason to minimise or eliminate the role of luck in other aspects of our existence. Therefore, the charge that accepting a role for luck is self-defeating appears false.

Let us turn, then, to the second objection. Wolf's account of why we ought to take responsibility (in some sense) for outcomes beyond our control is not spelled out in detail, and she does not give a clear or compelling reason for why we should accept such responsibility. My aim in what follows is to provide such a reason.

Let us try to lay Wolf's argument out formally:

Argument 1

P1. We exist in the actual world.
P2. In the actual world, luck[98] plays a role in how our actions affect others, and how the actions of others affect us.
P3. We should accept what is true.
C1. Therefore, we should accept that luck plays a role in how our actions affect others, and how the actions of others affect us.

[98] I am using luck here in the sense of "circumstances outside of our control." As is the case with the usage of "luck" in the context of moral luck, this sense of luck is somewhat atypical. I agree with Hartman (2017) that the fact that luck is being used in an atypical sense is not relevant to the problem, however.

P1 and P2 are uncontroversial. The conclusion follows from the premises, so the argument is both valid and sound. The problem is that the conclusion gives us no reason to think that accepting this role for luck means we should take responsibility for how — due to luck — our actions affect others. All it does is to establish that we should acknowledge the truth of the proposition that luck affects how things turn out.

Perhaps, however, the argument can be amended to overcome this difficulty. Just prior to the passage I quoted above, Wolf writes the following:

> The reason for objecting quite generally to an attitude of greater detachment and for commending an embrace of at least some of what lies beyond the sphere of one's will has less to do with a benevolent concern for others than with a view about what, for lack of a better word, might be called psychic health.
>
> (2001, p. 6)

It is not entirely clear what Wolf means by "psychic health," but it is plausible that it equates to something like "well-being." If this is right, then Wolf appears to argue that "embracing" (or taking responsibility for) what "lies beyond the sphere" of our will enhances wellbeing, and that this can ground the claim that we should accept RTR.

Justifying a moral requirement in this way could be somewhat problematic, given that Wolf writes that this reason (that is, the benefits to one's "psychic health") is not "clearly or narrowly moral" and has "less to do with a benevolent concern for others… ." (2001, p. 6) In other words, it seems as if Wolf holds that cultivating the nameless virtue and taking responsibility for what one is causally responsible for is merely prudentially beneficial. But if this is the case, it is unclear how RTR can be a moral requirement. If it is simply prudentially beneficial, why should disinterested third parties care whether agents act in accordance with it?

My response to this has two parts. First, recall that Wolf holds only that part of the justification for RTR has to do with one's psychic health. The other part of it relates to its kinship with generous action and the benefits such action brings to others. Therefore, part of the justification for RTR is clearly moral. Second, simply because something is prudentially good does not mean it is not also morally good.[99] Arguably, one's psychic health is both a prudential and a moral good.

Here is a second version of an argument in favour of taking RTR to be a true moral requirement, which incorporates the above considerations:

---

[99] Some philosophers hold that a moral good is one that is good for non-selfish reasons. See, for instance, Huemer (2014.) But a prudential good could also be a moral good if we had an obligation to ourselves to pursue a particular prudential good. Kant holds this view (*MM* 6:417-418.) For a recent book-length treatment of the idea that we can have duties to ourselves, see Schofield (2021.)

Argument 2

P1. We exist in the actual world.
P2. In the actual world, luck plays a role in how our actions affect others, and how the actions of others affect us.
P3. Taking responsibility for how our actions affect others (even when we are not morally responsible) would lead to existences with greater wellbeing, on average.
P4. Existences with greater wellbeing on average are better than existences with less wellbeing on average.
C1. Therefore, it is good to take responsibility for how our actions affect others (even when we are not morally responsible.)
P5. If something is morally good, then we should act to promote it unless it is overridden by something of greater moral worth.
C2. Therefore, unless it is overridden by something of greater moral worth, we should take responsibility for how our actions affect others (even when we are not morally responsible.)

Argument 2 avoids the problem of Argument 1, in that its conclusion clearly relates to the need to take responsibility, and P1, P2, P4, and P5 all seem unproblematic. The conclusions follow from the premises, and so that leaves P3. Do we have reasons to accept P3?

It seems to me that we do. Let's say that I am in a crowded hallway and accidentally bump against someone, causing her to spill her coffee on herself. This occurs not due to any negligence; I am paying an appropriate degree of attention to my surroundings and am not under the influence of any substances which might impair my coordination or awareness. It is simply due to the nature of crowded hallways. Consider two possible ways in which I might respond to this. In response 1, I immediately stop and apologise for bumping into her, check to see that she is okay, offer to get a cloth to clean up any mess, and so on. In response 2, I simply walk off, and if she (or one of her friends) accosts me, respond as follows: "It's regrettable that my bumping into you caused you to spill coffee on yourself. But I didn't do anything morally wrong, and so I don't need to take any responsibility for the fact that you have coffee all over yourself. I hope someone is able to get you a cloth to clean up the mess, but I don't want to do it because I have other things I'd like to be doing." It certainly seems as if it is better to live in a world in which, when I am causally but not morally responsible for bad

consequences, people behave in accordance with response 1 rather than response 2. If people behave in accordance with response 1, it maximises the chances that victims of bad luck are provided with assistance to overcome the consequences of their bad luck (in the case of the woman in the situation above, that someone will check she is not hurt and assist her in cleaning up and so on); and it also fills an epistemic gap in both the victim and in others, such as witnesses to the event or those who might hear about the event. By acting in accordance with response 1, I reassure these parties that I did not intend any harm and that I am appropriately considerate of the interests of others.

It is this latter consideration that, it seems to me, provides the benefits to one's "psychic health" that Wolf claims. If one were to eschew acting in accordance with RTR, then one would go through life with tenuous and uneasy relationships with others. If when I bumped into somebody in the supermarket I simply walked off, or said, "I didn't mean to do that" and then left, that person — and others who witnessed or heard about the event — would be left uncertain as to whether my action was truly accidental, or whether I had really been paying an appropriate amount of attention to my surroundings, and so on; and this could affect the way these people interacted with me in the future. They would likely treat me with wariness or perhaps even suspicion; and to be treated as such certainly does not seem conducive to one's "psychic health."[100]

If the argument above is correct, it shows that there is a consequentialist justification for the view that one is morally required (at least in some circumstances) to take responsibility for what one is only causally responsible for.[101] However, one might counter that this is not the phenomenology of the moral belief RTR. That is, we do not calculate that taking responsibility for what we are merely causally responsible for will produce consequences, and then act accordingly. Rather, it simply seems to us that RTR is a moral truth. This leads to a simpler justification: that we have an intuitive basis for taking RTR to pick out a moral truth.[102] The intuition arises in situations like the one I sketched out

---

[100] David Miller (2005) offers a view along these lines, arguing that, in cases in which there is no clear morally culpable agent for harm caused, we assign responsibility to whoever is merely causally responsible due to a human need to assign obligations for repairing the harm to someone.

[101] Adam Smith seems to put forth a similar view to this when he writes that a person who kills someone "without the smallest degree of blamable negligence" nevertheless rightly feels himself required to render to the family "every good office which he can devise or they accept of, to atone for what has happened." (*Theory of Moral Sentiments* II.iii.3)

[102] David Enoch (2012) argues for a requirement to take responsibility for what we are not morally responsible for largely from an intuitive basis. However, he also offers brief arguments for the contention that there is a "duty to take responsibility." Much of Enoch's essay deals with the type of responsibility-taking that an agent performs when she takes responsibility for the actions of her son, or of her country; but he also deals with the type involving a blameless driver who hits a pedestrian. Such an agent, Enoch holds, has a duty to take responsibility because "we should take interest in the world, not just in our will or intention, and so we should think of ourselves as creatures whose intentions and bodily movements have effects in the world. So we should incorporate the consequences of our actions (some of them, that is, in some circumstances) …into our agency

above that the agent ought to take responsibility for the bad consequences that have arisen as a result of his actions. As I wrote in Chapter One (see 1.1), I take it that, if it seems that *p*, then there is at least some justification for believing that *p* unless there are good reasons to doubt that it. Good reasons include the fact that *p* conflicts with another intuition that more strongly coheres with other beliefs, or that it conflicts with inferentially derived beliefs (beliefs derived from other intuitions); or that there are reasons to suppose it is infected by bias. Let us examine whether either of these reasons provides us with grounds to doubt the intuition RTR.

Let us consider the first reason: that RTR clashes with other, more plausible intuitions. It could be argued that RTR clashes with the more abstract intuition that one cannot be required to take responsibility for what one is not morally responsible for. Given that I have argued that, in general, abstract intuitions should be granted more credibility than intuitions that emerge in particular situations, this could be taken as a reason for doubting RTR. However, I think that there are better reasons for doubting the intuition that one cannot be required to take responsibility for what one is morally responsible for. For instance, it seems relatively uncontroversial to hold that — at least in some circumstances — one can be required to take responsibility for the consequences of the actions of one's children. Similarly, it is plausible that one can take (at least some type of) responsibility for the actions of one's country.[103] If this is correct, then the intuition that one can only be required to take responsibility for what one is morally responsible for appears false, and RTR cannot be considered implausible on the basis of it.

I will turn now to the second possible reason for doubting the intuition that grounds RTR: that the intuition is the product of bias. In Chapter One, I identified four possible sources of bias: self-interest, culture, emotion, and evolution. It is not clear how the intuition could be biased due to self-interest,[104] and nor does it appear to be the product of cultural bias. There is perhaps a better case to be made that the intuition is warped by emotion. After all, situations in which we perceive that we have such a responsibility are likely to be emotional ones, in some cases highly so. Driver_1, for

---

by taking responsibility for them." (2021, p. 127) This explanation to me seems to suffer from the same vagueness as Wolf's. However, there is much that is valuable in Enoch's article, and I engage with his ideas in more detail in Chapter Eight.

[103] For instance, see Goetze (2021, 2022) who uses the term "moral entanglement" to characterise situations in which an agent is responsible for consequences without being *morally* responsible for them. I investigate the nature of taking responsibility for what one is not morally responsible for in detail in Chapter Eight.

[104] Sinnot-Armstrong (2006) argues that morality touches on our self-interest so pervasively that, if we take the possibility of self-interest to be a reason to doubt the veracity of a moral intuition, we are not justified in taking any moral intuitions as true without further justification. Although it is perhaps true that many lower-level moral intuitions touch on our self-interest (either directly or indirectly), this seems less true of formal intuitions.

instance, perceives RTR in the wake of running down a child, a highly emotionally charged event. Even the situation I sketched above, where an agent spills coffee on someone in the hallway, is likely to evoke some emotion in the agent. However, that the intuition is not infected with bias as the result of emotion is clear from the fact that uninvolved third parties — say, those who merely hear of the event, and have no connection to either of the people involved — also intuit RTR. That the intuition also arises when we consider hypothetical situations further emphasises this. Therefore, while the intuition might often be *accompanied* by an emotion, there are good reasons for thinking it is not the *product* of or biased by an emotion.

Could the intuition be the product of evolutionary bias? One can imagine an argument that it is. For instance, perhaps in the environment in which we evolved (and conceivably even now) taking responsibility for what one was only causally responsible for led to reductions in animosity and conflict, because it made clear to the harmed party that the person causally responsible for the harm did not have harmful intentions, and therefore need not be distrusted in the future. This could plausibly have led to greater nonzero-sum cooperation. This is similar to a point I made above, where I noted that taking responsibility for what one is causally responsible for plausibly has the effect of ensuring that the harm is repaired (at least as far as is possible). If no such moral principle were adhered to, then it is plausible that it would lead to fewer people who have been harmed getting the assistance in repairing the harm.

But it is not clear that such an argument gives us good reasons to doubt that RTR is an actual moral requirement. First, even if the above explanation is possible, it does not show that it is correct. It seems equally plausible that the reason we have the intuition RTR is that it picks out an objective moral truth. Second, even if the intuition is the product of evolution, the consequentialist argument that I made above still stands, and so the argument from evolutionary bias against the truth of RTR at best establishes that it is not a fundamental moral principle, but rather one that we can arrive at inferentially. As nothing in my argument rests on whether RTR is non-inferential or inferentially derived, the fact that it might have an inferential basis does not present a difficulty for my view.

Therefore, not only are there no good reasons to doubt the moral requirement that the nameless virtue gives rise to; there are consequentialist considerations in support of it. I conclude, then, that Wolf's claim that it is a moral requirement to take responsibility for what non-culpable harm one causes is justified.

Finally, it is worth noting that RTR must be a prima facie requirement rather than an absolute requirement. This is because, on the view of moral requirements I outlined in Chapter One, all moral requirements are prima facie; that is, they can be overridden by other moral requirements in a

situation in which there is a conflict. This gives rise to some interesting questions concerning RTR. For instance, it seems as if there could be a situation in which an agent faced a requirement to take responsibility for what she was causally responsible for and another moral requirement to do *x*, but was unable to fulfil both. If she judged that the requirement that she do *x* overrode the requirement RTR, then plausibly she should experience some sort of second-order agent-regret, and face a second-order requirement to take responsibility. I will set this issue aside for now, however, because I analyse the matter in detail in Chapter Eight, where I sketch out a full account of what it means to take responsibility for what one is not morally responsible for.

In this section, I have made the argument that we have good reasons for taking RTR to be a true moral requirement. In the following section, I move on to analysing the emotion of agent-regret.

5.2: An Account of Agent-Regret

Above, I have laid out Wolf's solution to the problem of resultant moral luck, and shown how that solution can be applied to the problem of unavoidable moral wrongdoing to deny that it is possible. I have also provided a short justification for accepting that RTR, the moral requirement related to the nameless virtue, picks out an actual moral requirement. In this section, I expand on the nature of agent-regret.

My argument has been that Wolf's ideas can account for the phenomenology of cases of unavoidable moral wrongdoing. Part of this argument is that the moral distress an agent experiences in a situation of unavoidable moral wrongdoing is agent-regret. Therefore, for this solution to work, I need to show that it is plausible that there is such an emotion as agent-regret; and also that this emotion can be appropriate or rational. Some philosophers argue that agent-regret is either simply guilt or that it is irrational; and if this is true then my solution to the problem of unavoidable moral wrongdoing becomes untenable. My aim in this section is to show that there are good reasons to believe that there is an emotion of moral distress whose cognitive component differs from that of guilt, and that such an emotion can be appropriate.

First, I provide an argument for considering as agent-regret the moral distress an agent experiences in cases of unavoidable moral wrongdoing, which I define as an emotion of self-directed moral distress with a cognitive component different from that of guilt. Following that, I address an objection from Moore, namely that counterfactuals suggest that the emotion is best understood as guilt after all. Finally, I tackle what I consider to be the most serious objection against taking the moral

distress to be agent-regret: that self-directed regret in such circumstances would be irrational. In addressing this objection, I make a novel argument about the cognitive component of agent-regret.

5.2.1: An Argument for Agent-Regret

In this section, I make an argument that we have reasons to consider the emotion of moral distress that we experience in situations of UMW and RML (for instance) to be distinct from guilt. That is, I hold that this moral distress has a cognitive component that distinguishes it from guilt.

Agent-regret, as Williams (1981) articulates it and as I understand it, is a self-directed moral emotion that picks out consequences for which we are causally responsible but not morally responsible. Therefore, to defeat the objection that we have no reason to posit the existence of agent-regret, I need to show that there is an emotion of moral distress that is distinct from guilt (that is, one whose cognitive component does not pick out wrongdoing on the part of the agent.)

The best way to examine whether we have reasons for believing that there is such an emotion is to focus on situations in which the agent has unambiguously not done wrong. It is with reference to such a scenario that Williams first laid out the concept: the "lorry driver who, through no fault of his, runs over a child." (Williams 1981, p. 28)[105] To distinguish this driver from Wolf's version, let us call him WD_1. Here is an argument for the existence of an emotion consisting of moral distress and a cognitive component that does not involve the ascription of guilt to the agent experiencing the emotion:

P1. WD_1 feels self-directed moral distress.
P2. This moral distress is appropriate.
P3. WD_1 and third-party observers correctly believe WD_1 has not committed wrongdoing.[106]
P4. Guilt is only appropriate for an agent when he has committed moral wrongdoing.
C1. Therefore, the moral distress WD_1 experiences cannot be guilt. (from 1-4)
C2. Therefore, there is an emotion of moral distress that is distinct from guilt and that it

[105] Scenarios such as this are rife in the literature. See Enoch and Marmor (2007), Hanna (2014), Sussman (2018) and Enoch (2019.)
[106] It is possible that in some situations of this nature the agent in WD_1's position might use the word "guilt" to describe his emotional state. Michael Zhao (2020) gives such an example, referring to the case of Darien Strauss, who hit and killed a classmate while driving and was found to be not culpable. However, it seems plausible that Strauss is using the word in a broader sense encompassing moral distress generally, and that agent-regret more precisely picks out the exact sentiment he means to convey.

can be appropriate for agents to experience. (from 2, 5)

The argument is valid, so to dispute the conclusion one needs to challenge one of the premises. P1 seems obviously true; and it is hard to imagine anyone mounting a serious argument to the effect that an agent in WD_1's situation has done wrong, so P3 appears impervious to challenge, too. That leaves P2 and P4.[107]

As I noted in Chapter One, P2 is a key part of the phenomenology of situations of UMW. It is also, as Wolf articulates, a part of the phenomenology of cases of RML. However, it could be argued that, *contra* Williams, the emotions of self-directed moral distress that an agent such as WD_1 experiences are inappropriate. Perhaps in situations in which an agent is causally responsible for negative consequences but bears no clear moral responsibility, any self-directed moral distress is errant.

I address this objection at length in the following subsection, and in doing so sketch out a novel basis for the rationality of agent-regret. For now, I will simply say that such an argument intuitively seems dubious. Reflecting upon how we ought to feel if we were in a situation like that of WD_1, it seems that self-directed moral distress *is* appropriate, and that it is correct to assert, as Williams does, that it would be "an insane concept of rationality" (1981, p. 29) that dictated that no such distress was warranted. The fact that those who actually have caused harm of the sort mentioned in Williams's hypothetical have felt their distress to be appropriate gives further credence to this intuition. For instance, in a *New Yorker* article widely cited in the academic literature,[108] Alice Gregory tells the story of (the pseudonymously named) Patricia, who hit a motorcyclist while driving. Patricia's reaction is as follows:

> "Yes, it was an accident, and in a certain sense we were both to blame, but, at the end of the day, I *hit* him, I took *his* life," she said. "No matter how much you want to dismiss it as an accident, I still feel responsible for it, and I am." She cried, "I hit him! Why does nobody understand this?"
>
> (Gregory 2017)

Despite the fact that no one was at fault for the tragedy, Patricia nevertheless experiences moral distress which she considers justified. This suggests that dismissing P2 as false is implausible.

I will turn now to examine the possibility that P4 is false. A form of this claim has been made

[107] Of course, there could be a variety of emotions of moral distress apart from guilt, with agent-regret being only one of them. Maring (2021) suggests "stoic determination" as another evaluative component of self-directed moral distress.

[108] See Zhao (2020) and Wojtowicz (2022.)

by Michael Zhao (2020). Zhao takes the position that guilt is appropriate even when the agent is not morally responsible for the negative consequences he produces. On Zhao's view, an agent experiences guilt when "some part of the self … [is] implicated in a bad state-of-affairs," where "the self" is understood as some aspect of the agent that he identifies with. (2020, p. 307) For instance, one might identify with his "role as parent or philosopher," but not with "facts about [his] family history or appearance." (2020, p. 308) Whenever some aspect of the former sort is implicated in the engendering of negative consequences, Zhao holds, guilt is appropriate. Because we identify with both our agential and causal capacities, guilt is appropriate in cases such as that of WD_1. The agent need not be morally responsible for producing the consequences for guilt to be appropriate. Agent-regret, Zhao contends, is therefore merely a form of guilt. If this is correct, then P4 is false and my argument for the existence of agent-regret as distinct from guilt fails.

But Zhao's argument is flawed. Zhao's reason for considering agent-regret to be a form of guilt is that it is "phenomenologically and motivationally identical" to guilt. (2020, p. 299) But this is false on both counts. The phenomenology of guilt involves the affective state plus the cognitive evaluation that one has done wrong. The phenomenology of agent-regret has a different cognitive evaluation — that one is causally responsible for a bad state of affairs. Motivationally, guilt and agent-regret also differ.[109] Zhao is correct that in both cases the agent is motivated to make amends for the consequences of his actions. However, in the case of guilt, the agent is also motivated to *act differently in the future.* This motivation to alter one's behaviour in the future cannot be a feature of the emotion of WD_1, because the agent has made no mistakes that she must take steps not to repeat. If this is correct, then agent-regret and guilt do have different phenomenologies, and Zhao's attempts to conflate the two fail.

One other argument Zhao makes against the existence of agent-regret is worth addressing: that agent-regret cannot account for other kinds of causal responsibility:

> [T]here are various kinds of causal responsibility in play in the examples above, two of which don't involve agency at all. In [these two types of examples], the subjects needn't have done anything voluntarily that brought about the bad outcome. Rather, they were causally responsible for it in some other way, either instrumentally or indirectly.
>
> (2020, p. 298)

[109] It seems as if the motivational component of an emotion is merely another aspect of its cognitive component, but whether or not this is true makes little difference to my argument.

To illustrate causal responsibility, Zhao uses the following story, which he calls *Rescuer*:

> A man was driving to work when his car unexpectedly hit an icy patch on a bridge and flipped over. Local firefighters were sent to pull him out from the wreckage. In the process of reaching him, one of the firefighters slipped and fell through a gap in the roadway, plummeting to his death.
>
> (2020, p. 291)

This is not agential responsibility, Zhao argues, because the man's agency is not implicated in the death of the firefighter. Unlike WD_1, he does not *do* anything to cause the death of another; he is simply sitting in his wrecked car. Agent-regret cannot explain self-directed moral distress in cases such as *Rescuer*, Zhao holds. Nevertheless, the man feels distress. Therefore, it is better to understand what is typically called agent-regret as a form of guilt.

One response to this is that, if there is no agency involved in the infliction of bad consequences, it gives us even less reason to think that the emotion could properly be termed guilt. But even if we put this aside, it is not clear that the man in *Rescuer* is *not* agentially responsible for the death of the firefighter. To see why, consider a somewhat altered version of the case of WD_1 (let us call this driver WD_2). Like WD_1, WD_2 is driving his truck blamelessly. He is following all the road rules, is not drunk or under the influence of drugs, got an appropriate amount of sleep the previous evening, and so on. However, WD_2 is driving his truck 20 years into the future, when self-driving vehicles have become commonplace. WD_2 has therefore flicked the switch to put the truck in autopilot (let us stipulate that autopilot is as least as safe as human-controlled mode, and so WD_2 is not open to moral censure for this action.) Unfortunately, a child runs out in front of WD_2's truck and is struck and killed.

In this case, WD_2 is also passive at the point of the tragedy — he is simply sitting in his truck as autopilot manoeuvres it through the streets. However, it would be odd to claim that he was not agentially responsible for the death of the child. After all, he chose to drive the truck, to put the autopilot on, and so on. The fact that it was on autopilot when the tragedy occurred appears irrelevant. It seems as if the man in *Rescuer* has a similar relationship to his tragedy as WD_2 to his. While it is true that he does not do anything to directly produce the death of the firefighter, it is his action of driving onto an icy patch that precipitates the bad consequences, in the same way that it is WD_2's action in putting the truck in autopilot that precipitates the bad consequences. Therefore, it appears as if indirect responsibility is sufficient to produce in the agent the emotion of agent-regret. If this is right, then Zhao's argument is unconvincing.

Some might argue that while the cognitive components in the cases of guilt and agent-regret

differ, they are nevertheless similar enough that we are warranted in using the word "guilt" to describe both. However, this turns the argument into a matter of mere semantics. Obviously people are free to use whatever word they like for things; but the question is whether guilt (as the word is typically used) and what I am calling agent-regret are different in ways relevant to the issues of unavoidable moral wrongdoing and resultant moral luck. I think it is clear that they are. As I have argued, this difference lies in their cognitive components. If this is right, then the argument over whether agent-regret is a form of guilt or not becomes a bit like arguments about whether a hotdog counts as a sandwich, or whether chess is a game. All that matters for my argument about unavoidable moral wrongdoing is that there are situations in which

1. An agent feels self-directed moral distress;
2. The distress is appropriate; and
3. The agent has done nothing wrong.

Once these three points are conceded, one has to concede the existence of what I am calling agent-regret, even if she wishes to call it guilt. And if what I am calling agent-regret exists, then one cannot object to Wolf's account — and therefore my application of her account to situations of unavoidable moral wrongdoing — on the basis of its non-existence.

5.2.2: Moore's Objection

Above, I made an argument for the existence of an emotion of self-directed moral distress with a cognitive component different from that of guilt, and considered a challenge to P4 of my argument. Now, I turn to address some arguments others have made against the existence of such an emotion. One such argument has been put forth by Michael Moore. Below, I consider his argument.

Moore argues that there is no evidence that agent-regret exists. In particular, he argues that the phenomenology of near-misses undermines the existence of agent-regret. Moore argues that, when an agent is guilty of some wrongdoing, near-misses

> produce feelings of relief and judgments of the form 'there but for the Grace of God go I'. Wolf would have to account for such feelings and judgments as being about agent-regret and the 'nameless virtue' in feeling it: the near-misser just missed, not guilt, but the regret of having his agency involved in causing a bad state of affairs. He has escaped, not greater blameworthiness, but rather, a kind of 'dirty hands' that it would be his aretaic duty to regret. He thus feels relief. Surely this is inadequate as a description of what the near-misser both feels, and should feel. On its face, his feelings are those of relief that his culpable carelessness

> is much less blameworthy than it could have been had it caused the harm it merely risked. On its face he judges himself to be less guilty than he easily could have been, feeling grateful that he has been lucky enough to get a second chance to do better before his moral ledger is ruined forever.
>
> (2009, p. 32)

Moore thus gives two accounts of relief, and claims that the second is more plausible. On this second view, an agent such as Driver_2 (who does not ensure his brakes are maintained but kills or injures no one as a result of this), when reading about Driver_1's accident the next day, feels great relief, and this feeling of relief only makes sense if, had it been he who had had an accident, he would have been more blameworthy, and therefore also guilty of greater wrongdoing.

But it is not clear that Moore's argument is correct. It seems plausible that the relief Driver_2 experiences is due not to the fact that he has done lesser wrong, but because of the following two reasons:

1. He does not have to take responsibility for the child's death (and therefore does not take on any obligations arising from the death); and
2. His wrongdoing was not discovered by others.

1 is plausible because it is often the case that we experience relief at escaping moral obligations. For instance, imagine a relative of mine invites me to a birthday party. I know from her previous birthday parties that they tend to be dull affairs, but I consider it one of my obligations as a friend to attend. When she falls (mildly) ill and cancels the party, I feel a sense of relief at not having to attend. It seems plausible that Driver_2's relief has (in part) the same basis.

The plausibility of 2 stems from the similarly widely recognised experience. When we commit a wrong and no one finds out, we feel great relief. This is not because we believe the fact that no one found out makes the otherwise wrong act somehow right, but because we are glad at avoiding the damage to our moral reputation. Imagine a student who has not studied for an important exam (perhaps because she has spent too much time partying recently.) Knowing that she must do well on the exam, she scrawls some notes on her arm to surreptitiously consult throughout the exam. She knows this is wrong but does it anyway. The proctors do not notice her consulting her notes during the exam, and upon leaving the exam hall she feels a great sense of relief (mingled, perhaps, with some

guilt.) The plausibility of this story shows that an agent's wrongdoing going undetected can produce relief.[110]

Therefore, while Moore might be correct that P1 could not alone explain $Driver_2$'s sense of relief, it seems plausible that 1 and 2 together can. If this is right, then Moore's argument does not provide us with good reasons to doubt that agent-regret is a separate emotion from guilt.

5.2.3: Agent-Regret as Irrational

Another agent-regret related objection concerns its supposed irrationality. As I have noted previously, in situations of UMW and RML we have the intuition that the moral distress an agent experiences is appropriate. If we had good reason to suppose that agent-regret was irrational (and therefore inappropriate), then this would suggest that the emotion agents experienced in such a situation was guilt, and that therefore both RML and UMW were possible. In this subsection, I lay out the objection from the irrationality of agent-regret and then explain why it is ultimately unconvincing.

The objection from irrationality goes like this. On the view of emotion I laid out in Chapter One (see 1.4.1), an emotion consists initially of a rough, affective, non-cognitive evaluation, and this is then refined into a propositional cognitive component. Traditionally, the cognitive component (or "constitutive thought") of agent-regret has been understood as, "how much better if I had done otherwise."[111] Wojtowicz succinctly captures the standard picture of the cognitive component of agent-regret as follows: "In feeling agent-regret, an agent assesses the result as regrettable, and she recognises her responsibility for the result in virtue of her action, thus she comes to see her action as regrettable." (2022, p. 75) In other words, the agent evaluates the result of her actions as in some way morally bad, and therefore regrets her choice to act in that manner. However, it is unclear how such an emotion could be appropriate. In the case of $Driver_1$, for instance, he no doubt experiences moral distress over his decision to neglect to perform appropriate maintenance on his truck. However, because his decision to neglect to perform such maintenance was wrong, this moral distress seems to be guilt. Can it be rational for him to experience self-directed moral distress over and above this guilt? If so, for what?

Williams's suggestion, endorsed by Scarre (2017), is that it is rational for him to experience

---

[110] This echoes a point by Brian Rosebury, who notes that [i]t is just bad luck (or good luck) for a person's happiness that certain facts that would affect his reputation if known remain unknown." (1995, p. 521) As Rosebury notes, however, this is not moral luck — the agent's degree of wrongdoing does not hinge on it — but rather just plain luck.

[111] See Williams (1981.)

regret over the fact that he took that particular route, or left that particular time. If $Driver_1$ had taken a slightly different route, or left just a few moments later, or made some other minute adjustment to his behaviour, that would have resulted in the tragedy being averted. Perhaps the regret is over the fact that he made these particular choices. However, it is not clear how such regret could be rational. It is rational to regret one's choices when one's choices do not accord with one's reasons for action. This is why it is rational for $Driver_1$ to experience guilt (a specific type of regret) with regard to his negligence. However, WD_1 had no reason to decide to leave later, or take a slightly different route, or to stop for a coffee in the middle of his journey. Therefore, it appears as if his agent-regret must be irrational.

This seems to be even clearer in the case of an agent in a situation of UMW. Consider Sophie. She does not wish that she had acted differently, say by refusing to nominate one of her children to live. If she had acted in this way, then both of her children would be dead. As this is worse than merely having one dead child, she has no reason to wish she had acted differently — her original action was perfectly rational. She would obviously wish that she had not found herself in such awful circumstances, but this is different from wishing that her own actions had been otherwise, and because this regret is directed at circumstances rather than the self it is not clear how an emotion with such a cognitive component could count as *agent*-regret. Therefore, it might be concluded, agent-regret must be irrational. Because it seems to be the case that the emotion an agent such as Sophie experiences is appropriate, it appears that the emotion she experiences is not agent-regret at all. If this is correct, then Wolf's account appears in trouble as a solution to the problem of resultant moral luck, and so also as a solution to the problem of unavoidable moral wrongdoing. Is there a way of responding to this charge that agent-regret must be irrational?

5.2.4: Possible Responses to the Objection from Irrationality

In this section, I lay out several possible responses to the argument that agent-regret is irrational. I show that none of them provides a convincing response to the argument. This sets up the following section, in which I give a novel account of the rationality of agent-regret.

One response is to say that it is appropriate for the agent to experience an emotion with a cognitive component along the lines of the following: "How much better if I had not produced feature *a*," where *a* is the morally bad consequence of the all-things-considered morally right action. On this view, Sophie's agent-regret is appropriate because, although she does not regret her actions all-

things-considered, she regrets the feature of the action that contravenes her values.[112] However, such a proposal is unconvincing, as the intentionality of this emotion is once again towards the circumstances that coerced one into producing the morally objectionable feature, not towards the action of the agent. This is because, unless the agent regrets the outcome of the action all-things-considered, she does not regret the action that produced the feature (or if she does, such regret is inappropriate.) Therefore, agent-regret cannot have a cognitive component along the lines suggested above.

Another possible argument for the rationality of agent-regret comes from David Sussman. Sussman argues that agent-regret is rational because of the fact that the agent and the person she harms (or might harm) have been thrust into what he calls a "moral state of nature" in which the "only way for each to defend their basic rights is to violate the rights of the other." (2018, p. 804) It is plausible, Sussman holds, "that a person is entitled to use what would otherwise be a wrongful degree of force in order to stop an innocent threat to their lives." (2018, p. 802) To illustrate how this can ground a claim about the rationality of agent-regret, he asks us to imagine a pedestrian who is blown by an unusually strong gust of wind onto the road and into the path of a truck. Neither the pedestrian nor the truck driver is guilty of any type of negligence or does anything wrong. In this situation, Sussman argues, the pedestrian would be permitted to use a disintegrator gun (if such guns happened to exist and if she were carrying one) on the truck to stop it from striking and killing her. This is because the truck presents a threat to her life, albeit an innocent one. For the same reason, the truck driver would be permitted to use his own disintegrator gun to prevent the pedestrian from disintegrating him first. The driver and the pedestrian can therefore find themselves in a situation in which each is entitled to kill the other.

Sussman holds that this is what makes agent-regret rational. Once a crisis such as the one articulated above is over, the relevant parties need some way of signalling their return to what Sussman calls normal moral relations. Sussman argues that

> [i]n the absence of any more familiar kinds of gestures that might do this work, our best alternative lies in recruiting familiar rituals of apology and reconciliation into service that go beyond their primary function of repairing relationships damaged by wrongdoing. The driver, like someone who is truly in the wrong, asks for forgiveness and tries to make amends as a way of acknowledging that the pedestrian was indeed entitled to treat him as an enemy.
>
> (2018, p. 805)

In other words, because we have no established methods of signifying our desire to reconcile after

[112] See Tannenbaum (2007) for an argument along these lines.

being in a moral state of nature, we use words and forms of behaviour from the domain of moral wrongdoing to perform the task. In apologising and seeking to take responsibility in whatever way he can, the driver demonstrates the return to normal moral relations, and the pedestrian makes a similar demonstration by accepting his apology and forgiving him. Agent-regret is therefore rational, on Sussman's view, because it is rational for the agents to regret being put into a position (through bad luck) in which they stand in a morally antagonistic relationship to each other.

But there are some significant problems with Sussman's solution. The first problem has to do with the cognitive component of the emotion of agent-regret. On Sussman's view, agent-regret arises when, through bad luck, an agent is in a situation in which it is permissible for her to violate the rights of another agent. Therefore, the cognitive component of the emotion must be something like, "How much better if I were not in this position." However, if this is the cognitive component, then Sussman's account is subject to the same objections as some of the other accounts above in that it is not clear that an emotion with such a cognitive component is self-directed. That is, it seems as if the regret is directed at the circumstances that produced the situation, and not at oneself. If this is correct, then Sussman's view has the same problematic features as accounts such as Scarre's in that it cannot account for the self-directed quality of agent-regret. It is therefore unclear why we should accept his account.

Second, it is not apparent why it is only the truck driver for whom it is rational to experience moral distress, and not both parties. If as Sussman claims it is rational to experience moral distress at being thrust into a state of moral nature with another agent — a state in which it is permissible to violate the other agent's rights — then both the driver and the pedestrian from Sussman's example would seem to be in the same position. Why, then, is it only rational for the driver to feel a type of moral distress which prompts him to adopt a submissive and apologetic posture? Why should the pedestrian — who was also, on Sussman's view, part of the moral state of nature, and ready to use her disintegrator on the driver — not experience a similar emotion, and adopt a similar posture?

Sussman addresses this objection briefly in a footnote, commenting that "all the injuries are borne by one party, a significant fact that we can register by giving the role of the apologizer to the person who had more power (and less luck) in the situation." (2018, p. 805 footnote)[113] In other words, truck driver's experiencing agent-regret and the pedestrian's adopting the role of wronged victim are a way of making up for the unfairness of the situation. But this is not the phenomenology of agent-regret. The truck driver does not perceive himself as doing the pedestrian some sort of

[113] What Sussman means here is obviously that the pedestrian has less *good* luck. The fact that the truck driver finds someone in the path of his truck and the fact that the pedestrian finds himself in the path of a truck seem lucky to the same extent.

favour by taking on the role of apologiser; and nor does the pedestrian experience the apologies he receives as the mere enactment of a social ritual. Rather, the truck driver experiences a sense of self-directed moral distress which prompts him to behave in such a way as to ameliorate the damage his actions have caused; and the pedestrian experiences a feeling that he is the harmed party and deserving of reparations. But Sussman's account cannot explain why either the driver or the pedestrian would have these emotions, because on his account both agents are simply enacting a social ritual.

Even if the above argument is not correct, there is a further reason to find Sussman's solution unsatisfactory. It is that it leads to the implausible view that in near-miss situations both agents should feel agent-regret. Imagine that instead of hitting the pedestrian, the truck driver manages to brake just in time, screeching to a halt just a couple of inches in front of her. In this situation, there is no injured party, and yet on Sussman's view both parties (prior to the truck coming to a stand-still) would still have been in a state of moral nature and justified in drawing their disintegrator guns and using them on each other. As there is no party who has worse luck than the other, there is no reason to distinguish between the two agents, and so presumably both should signify the end of their personal antagonism by apologising and experiencing agent-regret. However, this is absurd. It is easy to imagine the truck driver feeling immense relief at not having hit the pedestrian; and certainly the pedestrian would feel (even greater) relief at not having been hit. But it is utterly inconceivable to imagine either of them experiencing anything like agent-regret. What this shows is that it is not a situation of personal antagonism arising from being in a moral state of nature that generates feelings of agent-regret. Rather, it is the bad consequences that trigger the emotion. Sussman's view is therefore unsuccessful in accounting for the rationality of agent-regret.

Another account of the rationality of agent-regret comes from Jake Wojtowicz, who proposes that agent-regret is rational because what we regret is not the decision that produced the regrettable outcome, but the action we take, and our (causal) responsibility for the outcome. One can endorse the decision to perform action *y,* and yet "regret the exercise of agency that left [one] responsible, as an agent, for that outcome." (2019, p. 64) This is possible, Wojtowicz holds, because assessment of a "decision itself turns around its intrinsic qualities [that is, whether one made the best choice given the information that was available to her]," whereas an "action can be made good or bad in light of what follows. (2019, p. 64) When one experiences agent-regret, "what follows" the action is bad, and so one is causally responsible for a bad outcome. However, divorcing endorsement of a decision from endorsement of the action that follows from the decision is implausible. Plausibly, when one decides to perform an action, one believes that one has the most reasons to perform that action. If, later, one

comes to regret performing the action but does not regret deciding to perform the action, then one is regretting acting in accordance with what she had the most reason to do. This is irrational.

Wojtowicz attempts to counter such a line of argument by broadening the concept of an action:

> [A]n action is made of more than just a decision, it can be made of a result, too; and our actions can also be significant in terms of their consequences. It is the fact I am responsible for your broken leg that I regret, but this does not mean I have made any sort of error in my decision that needs to be renounced, and I can fully endorse deciding that way.
>
> (2019, p. 64)

Here, Wojtowicz suggests that the term "action" should be understood (at least in the context of the agent-regret debate) to include the consequences of the action. But simply runs into the same problems. If I endorse a decision procedure, then rationally I must endorse the action that accords with the conclusion of the decision procedure. But, as I showed in the argument above, if I endorse the action it is irrational for me to regret the fact that I am responsible for the negative outcome.

Another possible response to the objection is through adapting another Wojtowicz idea: the concept of "pure agent-regret." As Wojtowicz conceives it, pure agent-regret is when "one regrets what one has done yet this cannot be explained by one's independent regret (e.g. where one would feel ordinary regret were one a bystander) about the result." (2022, p. 76) To illustrate the concept of pure agent-regret, Wojtowicz asks us to imagine that Driver_1 is Bob Dylan, and instead of running over a child he runs over one of the warmongers whom he rails against in his song "Masters of War."[114] Given Dylan's views of the warmonger whom he runs over, it is implausible to suppose that he regrets the outcome. However, it is equally implausible to suppose that he feels no agent-regret at the fact that *he* is the one who killed the warmonger. Therefore (Wojtowicz's argument goes), agent-regret is at least on some occasions "pure" in the sense that it is not the product of the consequences of what one has done, but simply regret that one's agency was involved in producing the consequences. Agent-regret, on this view, involves regret at how our actions "affect our identities in important ways." (Wojtowicz 2022, p. 84) In the example above, Bob Dylan does not regret the death of the warmonger, but he does regret the fact that his identity has changed: he has become a killer. If this is correct, then it does not matter that the agent does not regret the overall outcome of his actions. What matters is that, in producing an outcome that he endorses, the agent nevertheless regrets that his identity has been altered in a way he considers bad.

[114] The lines of this song include the following sentiments, addressed to the eponymous figures: "And I hope that you die/ And your death'll come soon." (Dylan, 1963)

Wojtowicz's explanation for how agents such as (the fictional version of) Bob Dylan can experience agent-regret, despite not regretting the consequences, can be adapted so as to explain how agents in situations of unavoidable moral wrongdoing and RML can appropriately experience agent-regret. Consider his comments on the actions of a Dutch woman who was part of the resistance against the Nazis. Drawing on Gaita (2004), Wojtowicz describes her case as follows:

> [She was] in a plot to assassinate [Hitler] and had to turn three Jews out of her protection so as not to risk the plot (they were killed). Although she clearly regrets the death of those she was protecting, one can also recognise a self-focussed element. It is not just what happens that matters, but what performing these actions does to somebody's own sense of who they are. She is now, in her own eyes, a murderess.
>
> (2022, p. 83)

Wojtowicz argues that the fact that the Dutch woman's identity has been altered through her actions, (partially) explains her feeling of agent-regret. Despite not committing a morally wrong act (*ex hypothesi*), she is now a killer. As being a killer is something to be regretted, it is perfectly rational that this should be a source of regret to the woman; and unlike the proposals above, this regret is also self-directed. Therefore, this appears to provide a plausible response to the objection.

Wojtowicz's suggestion about identity seems to generalise. Consider a case from the Clint Eastwood film *Million Dollar Baby*. In the film, Frankie is an ageing boxing trainer. One day, Maggie (played by Hillary Swank), a struggling waitress who has never boxed before, comes to ask Frankie to train her. Although he does not believe that women should box and so at first refuses, he comes to respect Maggie and so changes his mind. Her determination, coupled with Frankie's instruction, leads her to improve rapidly, and she quickly earns the right to fight for the world title. In the title fight, however, she is knocked down by a blow from her opponent and paralysed from the neck down. Seeing her life as having no purpose, Maggie asks Frankie to help her commit suicide. He initially refuses but then comes to realise that it is the right thing to do, and euthanises her by lethal injection. The last shot we get of Frankie in the film shows him shrouded in darkness as he exits the hospital. The film clearly means us to see that Frankie does the all-things-considered morally right thing in helping Maggie end her life. However, it shapes his identity in ways it is perfectly rational for him to regret. As Frankie exits the hospital, his assistant, Scrap (played by Morgan Freeman) comments, "I don't think he had anything left," and says that we learn that Frankie never returns to the gym, and speculates that "he [doesn't] have anything left in his heart." (Eastwood 2004) Furthermore, slightly earlier in the film, when Frankie is contemplating whether to accede to Maggie's request, he confers with his priest, who tells him, "If you do this, you'll be lost… somewhere so deep, you'll never find yourself again." What this suggests is that what one does — even when it is morally right — shapes

one, and not necessarily in ways that one likes. Despite doing what is right, Frankie nevertheless has to live the rest of his life with the memory of injecting Maggie with a lethal drug and witnessing her final breaths. What the film seems to suggest, then, is that even when it would be irrational for one to regret an action, one can nevertheless experience regret over the effects the action has on one's identity. As the case of Frankie shows, an all-things-considered morally right act can nevertheless produce alterations in one's identity that are far from welcome.

It therefore seems plausible that all-things-considered morally correct actions can nevertheless shape the identity of the agent in a negative way. However, should they? It is unclear why acting in an all-things-considered morally correct manner should negatively affect one's identity. Wojtowicz justifies this negative effect as follows:

> Now, one might worry that to see being a killer as a bad thing, as something that might make one's action regrettable, one must see the results of one's killing as a bad thing. But I think that would be a mistake. To adapt an example from *Les Misérable*s, Jean Valjean might think that being a thief is a shameful thing, even though he does not think there was anything shameful in his particular case where taking the loaf of bread was necessary to feed his starving sister. (2022, p. 84)

Wojtowicz's view seems to be that when the behaviour that one engages in would *typically* be wrong, that behaviour exerts a negative effect on one's identity even when it is *in this specific instance* justified. But it is not clear why this would be. In the example above, for instance, surely the reason that Valjean considers being a thief to be a shameful thing is that it is, under normal circumstances, morally wrong. But his particular case of thievery — stealing a loaf of bread in order to feed his sister — he does not consider to be an act of wrongdoing. It is understandable for one in Valjean's position to feel that his identity has been in some way altered for the worse; but it is not clear why it would be rational for him to think this. Therefore, despite its initial plausibility, the identity solution to his problem also seems to fail.

Finally, Jordan Mackenzie (2017) has also recently put forth an account of the rationality of agent-regret. Mackenzie argues that agent-regret is rational in the sense that it is part of a social practice. An emotion, Mackenzie holds, can be rational if it is a part of a larger social practice that is itself rationally justified. Mackenzie illustrates this through analogising it with the emotion of embarrassment. Embarrassment is rational, Mackenzie claims, because it is a breach of the social practice of etiquette. One would feel embarrassed to turn up at a wedding without one's formal wear — even if the lack of appropriate clothing were outside of one's control — because it is a breach of the larger social practice of etiquette. Similarly, because "the social practice of moral luck" is rationally justified — because it helps us "navigate the tragedies that result from bad moral luck," —

so too is agent-regret internally justified.

But the analogy between agent-regret and embarrassment breaks down. It is true that embarrassment arises when an agent breaches some aspect of etiquette, and so the emotion is rational within a set of social practices. However, the cognitive component of embarrassment is something like, "I wish [not *x*]," where *x* denotes the particular etiquette-transgressing behaviour. There is nothing at all irrational about experiencing an emotion with this cognitive component. However, the same is not true of agent-regret. Even if, as Mackenzie claims, we have a social practice of moral luck, and this social practice is rationally justified, it is unclear how this fact can mean that an agent who, *ex hypothesi*, has nothing to regret, can rationally experience an emotion whose cognitive component is "I wish I had done otherwise."

Nevertheless, in suggesting that agent-regret is rational in a second order sense, Mackenzie is onto something. I elaborate on this in the next section, where I lay out my own account.

5.2.5: Agent-Regret and the Requirement to Take Responsibility

In the previous section, I looked at several possible counters to the objection that it would be irrational to experience agent-regret. I concluded that none of them was successful. In this section, I aim to sketch out a novel account of the rationality of agent-regret. This account requires recognising that an emotion can be rational in a strategic sense; and that this strategic rationality depends upon the role agent-regret plays in alerting the agent to the demands of morality.

I think the key to showing that agent-regret is rational lies in the solution to the normative version of the paradox of fiction.[115] The normative version of the paradox of fiction is stated by Kim as follows:

1. We feel a rational emotion toward someone or something only if we believe that the object of our emotion exists.
2. We know characters and situations in fiction do not exist.
3. We feel a rational emotion toward fictional characters or situations.

(2010, p. 107)

[115] The paradox of fiction was first identified by Radford (1975). Dos Santos (2017) notes that there are two versions of this puzzle, the "possibility puzzle" and the "rationality puzzle." The former is about *how* it is possible to experience an emotion when one does not believe the proposition that is the object of the emotion.- The latter is normative, and is the version I focus on here. It accepts that we experience an emotion, and instead considers whether or not this emotion can be rational. For similar formulations of the normative paradox as a triad of inconsistent propositions, see Joyce (2000 p. 210), Matravers (2005), Gendler and Kovakovich (2011), and Teroni (2019.)

The paradox lies in the fact that all three of these propositions seem true, and yet taken together they are inconsistent. Let us consider this inconsistency with reference to the emotion of fear. Fear consists of the cognitive component, "I am in danger," or "someone I care about is in danger," along with negative affect directed at whatever the source of this danger is (an oncoming car, a drooling, barking dog, etc.) However, we also experience fear in situations in which we are not in danger, and nor is anybody or anything we care about. This is evident when we consider our emotional responses to works of art. For instance, let's say I am watching the movie *Halloween,* and see the murderous psychopath Michael Myers loom behind Laurie Strode (the heroine), pull out and brandish a large knife, and stealthily approach her, clearly intent on inflicting violence on her person. I am likely to respond by experiencing fear, but I am in no danger. It is not as if Michael Myers can walk through the television screen and continue his murderous rampage in my living room. Nor is anyone I care about in danger. I may care about the character of Laurie Strode; but Laurie Strode is not a real person. She is simply Jamie Lee Curtis reciting (in a convincing fashion) a series of lines from a script. She cannot die or be harmed; and nor can anyone else in the film. Therefore, the emotion appears to be (to use Patricia Greenspan's phrase) "representationally irrational."[116] And yet, despite this irrationality, when one experiences fear upon viewing *Halloween*, we do not dismiss it as silly. Indeed, we would consider it highly peculiar if one were to sit through *Halloween* and dismiss all fear in this way.[117]

If emotions towards works of fiction are irrational, what, then, explains the fact that we have the intuition that such emotions are rational? It seems to me that we judge it to be rational in a strategic sense. When we watch a good movie or a play, or read a good novel, we derive enjoyment or edification from identifying with the characters and the situations they find themselves in, and with their fears and aspirations, their triumphs and struggles. In other words, we find it beneficial to refrain from reminding ourselves that the characters are not real people, and from attempting to override our emotional reactions to them. If we thought to ourselves that such emotions were irrational, because the characters were not real people but simply words on a page, or actors reciting lines; and then strove to cultivate the disposition to not experience any emotions when consuming works of art, it would cease to make sense to consume any works of art. Reading a novel or watching

---

[116] Greenspan notes the distinction between "the rationality of the emotional state itself as opposed to the rationality of either inducing or expressing it." (2000, p. 478) The former she calls representational rationality, and the latter instrumental rationality. See also D'Arms and Jacobson (2000) for more on the distinction between an emotion's representational rationality (or "fittingness") and its instrumental or "strategic" rationality.

[117] For arguments in favour of the view that emotions directed towards works of fiction are irrational — the irrationalist position — see Radford (1995) and Matravers (2005.)

a movie would be pointless.[118] Because we recognise the consumption of works of art as a good,[119] it makes sense not to dismiss the emotions we experience in the course of this consumption. In other words, such emotions are rational for strategic reasons.

I will comment briefly on why it is that the consumption of works of art — and experiencing emotions in response to them — is a good. I largely endorse the view that it is due to the fact that it increases our ability to act rationally in the real world. For example, Gendler and Kovakovich write as follows:

> By engaging emotionally with fictional characters and situations, we broaden our range of simulated encounters, gaining insights about others' experiences that are processed much as if they had been our own. Without such a capacity, actual experience would be our only source of such emotional encounters, severely limiting the range of our reactive possibilities. So fictional emotions may contribute to our capacity for rational action through the role they play in educating our sensibilities. If so, then there is little reason to think that we should endorse a categorical principle according to which we would, ideally, fail to feel such emotions.
>
> (2011, p. 235)

It seems to me to be correct that our emotional responses to fictional characters can perform this function.[120] However, Matravers suggests that even if this is true, it does not make the emotions rational. This is because we do not experience emotional responses to fiction in order to increase our practical rationality, or to enhance our moral development:

> …[I]t is plain that our actions in reading a novel (and thus generating these emotions) are not guided by our subjective reasons… [A]lthough our reading of novels could be guided by some desire for self-improvement with respect to our knowledge of our emotional selves, that would be quite exceptional… Our motive for engaging with fictions … is the intrinsic rather than the instrumental value it brings.
>
> (2005, p. 260)

Matravers therefore contends that we do not engage with fiction to achieve some other goal, but purely for the sake of enjoyment. But there are a couple of problems with this line of argument. First,

---

[118] It is possible that certain forms of abstract art, or types of theatre such as Epic Theatre, could still be meaningfully engaged with under these circumstances.

[119] I mean "good" here in the prudential sense, although it is of course plausible that responding emotionally to works of art is a moral good, as well.

[120] Joyce offers a similar view, commenting that emotional responses to fiction "give [one] life experience on the cheap," teaching one important lessons which one can recognise the truth of and internalise. (2000, p. 223). See also Martha Nussbaum, who writes that "[a]llowing oneself to be in some sense passive and malleable, open to new and sometimes mysterious influences, is a part of the transaction and a part of [reading's] value. Reading novels, as David Copperfield learned, is a practice for falling in love. And it is in part because novels prepare the reader for love that they make the valuable contribution they do to society and to moral development." (1990, p. 238) Cf. Robinson (2005, pp. 146-148.)

it is not clear that it is true that our action in consuming a work of art is not guided by the desire for self-improvement. For instance, if I watch an Ingmar Bergman film, it seems entirely plausible that I am doing so (at least in part) in order to educate my moral sensibilities or improve my ability to understand and therefore to navigate the world. Second, even if I consume a work of popular entertainment — *Mission: Impossible — The Final Reckoning*, say — the fact that I expect I will enhance my well-being by experiencing a thrill when I see Tom Cruise leaping onto the wing of an aeroplane in midair or foiling an evil scheme *in itself* provides me with reason to experience the emotion. In other words, the fact that I judge that experiencing the emotions has intrinsic value is enough to ground their rationality in a strategic sense. In this sense, then, I depart from the justification for the rationality of emotional responses to fiction articulated by Gendler and Kovarich above.

Having made the case that emotional responses to fiction can be rational, I now turn to make my argument for the rationality of agent-regret. It appears plausible that we judge agent-regret to be rational for similar reasons.[121] On this view, although it is irrational to experience an emotion whose cognitive component involves regretting the result of an action that one had no rational basis not to take, we judge that it is appropriate or rational to *act as if* the emotion is rational; and for this reason it is rational, in a second-order sense. It is rational to act as if this emotion is rational because the consequences of accepting it as rational are better than the consequences of dismissing the emotion as irrational. This is why, as I commented above, there is a grain of truth to Mackenzie's and Sussman's analysis of the rationality of agent-regret. Recall that Sussman (2018) argues that agent-regret is rational because it allows the two agents to repair the wounded moral relations that have arisen as a result of being in what he calls a moral state of nature. I reject Sussman's view that the agents are in a moral state of nature, or that even if they are this can rationally ground the cognitive component of agent-regret. But his suggestion that the rationality of agent-regret stems from some strategic consideration seems to me to be correct.[122]

Note that I am not arguing that we reason that it would be good to experience some emotion, and then deceive ourselves into somehow experiencing the emotion. We do not, when viewing *Halloween,* decide that it would be good to experience some fear, and strive to trick ourselves into believing that Michael Myers really exists and could butcher us or our friends and family. Similarly, we do not recognise that we are causally responsible for harm and then strive to experience some moral distress. This is obviously at odds with the phenomenology of both situations. When we

[121] Jacobson (2012) suggests that, although agent-regret is irrational, it is admirable. This differs from my proposal in that I am not arguing that agent-regret is irrational, but rather that it is rational in a strategic sense.
[122] Sussman does not spell this suggestion out, but it is implicit in his account.

experience agent-regret, for instance, the emotion arises spontaneously, without conscious control. What I am arguing, then, is that we experience the emotion, and then perceive that we have strategic reasons not to dismiss it as irrational. Why do we experience the emotion of fear when watching *Halloween,* or moral distress when causally responsible for harm? The reason is that, in both cases, the emotions arise in situations which provide certain grounds for these emotions.[123] Typically, the sight of a knife wielding killer would make fear rational, and so the immediate response is warranted. Similarly, causing harm often warrants the emotion of guilt, and so the immediate affective response is understandable. Neither of these emotions is representationally rational in an all-things-considered sense; if we consider that Michael Myers is not real, the fear seems irrational, and if we recall that we are not morally blameworthy for the bad consequences, the moral distress seems irrational. But there is some evidence that grounds the emotional response.

Why is it strategically rational to accept agent-regret? The reason unites agent-regret with the nameless virtue and the moral requirement it grounds: the requirement to take responsibility for what one is merely causally responsible for. As I have argued (see 5.1), there are two plausible justifications for taking this to be a moral requirement, one intuitive and one consequential. Assuming that at least one of these justifications is correct, then it is true that we have such a responsibility. If we have such a responsibility, then it makes sense to respond with moral distress to situations in which we are causally responsible for negative consequences. This is because this moral distress draws our attention to the fact that we are causally responsible for negative consequences, and motivates us to take action. If we reject agent-regret as rational because of its reflexive intentionality (that is, the fact that it is self-directed) then what we are left with is regular regret. But regular regret does not motivate one to take responsibility for negative consequences; as Wolf (2001) points out, we would be somewhat puzzled by one who witnessed an agent accidentally break a vase, and then rushed over to offer to pay for it herself. Agent-regret, however, *can* provide this motivation. It therefore helps agents to overcome weakness of will and to act on the (all-things-considered) correct course of action.[124]

Should we not, on the view I have sketched out here, simply call agent-regret guilt? After all,

---

[123] This has certain similarities to Greenspan's notion that emotions can be rational perspectivally: "To say that an emotion is reasonable, or rationally appropriate, is to say that a certain evaluative belief that represents the content of the emotion (for anger, for instance, that someone has done me a wrong) would be warranted by a significant subset of the evidence… ." (2004, p. 218)

[124] Enoch offers a justification of agent-regret along these lines: "[there is a] moral requirement to take responsibility that the agent violates. Feeling agent-regret, I want to suggest, may be appropriate in that it is a way of acknowledging the existence of this duty. " (2012, p. 130) But while Enoch notes a link between agent-regret and the requirement to take responsibility, he does not explain how it is rational to experience an emotion whose cognitive component involves the thought that it would be better if one had done otherwise.

I have argued that the emotion arises in response to an agent bringing about a bad outcome. Such an emotional response is representationally rational in cases when an agent is morally responsible for the bad consequences; but when we are not morally responsible, its rationality is strategic. But if the emotional response is rational in both cases, is the emotion not the same, too? This line of argument appears to be strengthened through examining the example of fear I gave above. We do not have a separate name for the emotion we experience when we view Michael Myers on the point of committing his atrocities.

It is true that we do not have a separate name for this emotion; but it does not seem true that the emotion is entirely the same. For instance, when I am watching *Halloween* and experience fear at the appearance of Michael Myers wielding a knife, I do not perform the actions I would perform were Michael Myers to appear in my living room. I do not, for instance, rush off to barricade myself in my room, call the police, or grab a hockey stick for self-defence.[125] This suggests that the fact that the emotion is strategically rational is shaping the cognitive evaluation in some way.[126] It appears that the same is true of the moral distress an agent experiences when she is merely causally responsible for bad outcomes. In these cases the agent performs some of the actions that she would perform if she were morally responsible (apologising and attempting to make amends somehow, for instance.) But there are certain actions that she does not consider it appropriate to perform (turning herself in to the police, altering her behaviour in the future, etc.) Therefore, it is plausible that the cognitive component in the case of agent-regret is shaped by the fact that the emotion is rational in a strategic rather than representational sense.[127]

It should be noted that I am not arguing that experiencing agent-regret is necessary in order to perceive that there is a moral requirement to take responsibility for what one is causally responsible for. While experiencing agent-regret draws an agent's attention to the harmful consequences, and therefore to the fact that this moral requirement may apply, the requirement itself is derivable rationally. As I argued above (see 5.1), it is plausible that the requirement is intuitively derived; and even if this is incorrect, it seems as if one may arrive at it inferentially. Were Data from *Star Trek* to materialise on Earth, for instance, he would be able to perceive that there was a moral

---

[125] See Joyce (2000, p. 215) and Walton (1978, p. 7.)

[126] It has been suggested that, because emotions towards works of fiction do not motivate the agent to take action typically associated with emotion, they are quasi-emotions rather than emotions proper. As Walton puts it, "[f]ear emasculated by subtracting its distinctive motivational force [that is, to flee or to take action to avert the danger] is not fear at all." (1990, p. 202) However, as Matravers and others have pointed out, we are not motivated to take the action typically associated with an emotion in the case of emotions towards historical figures, either; and yet it seems strange to consider these quasi-emotions. (Matravers 2005)

[127] For more on how agent-regret motivates action, see 6.3.

requirement to take responsibility for the consequences of one's actions, despite his inability to experience emotions. As I noted above, the emotion draws attention to the moral requirement, and increases the agent's motivation to act upon it.

My argument here has been as follows. It is plausible that the emotions we experience towards fiction are rational in a strategic sense. I have argued here that the moral distress agents experience when they are causally responsible for negative consequences arises as an automatic response to circumstances which would often render the moral distress representationally rational. The reason we do not override the distress is because it is rational in a second order sense, a strategic sense. It draws attention to the negative consequences for which we are causally responsible and motivates us to act in accordance with the nameless virtue and the moral requirement associated with it: the requirement to take responsibility for what we are merely casually responsible for.

Conclusion

In the previous chapter, I argued that Susan Wolf's solution to the problem of RML could plausibly be adapted to show that UMW was not possible, and termed this the Nameless Virtue (NV) View. In this chapter, I have greatly elaborated on two concepts critical to Wolf's solution: agent-regret and the nameless virtue. In doing so, I have filled a gap in Wolf's proposed solution. Initially, I argued that we are justified in holding that the nameless virtue grounds a moral requirement to take responsibility for what we are merely causally responsible for. One justification for holding this is consequentialist; it seems as if the consequences of taking such responsibility are better than the consequences of not taking such responsibility. A second justification is intuitive. It simply seems that this is an actual moral requirement, and there are no clear defeaters for the intuition. On the normative ethical view I outlined in Chapter One, then, we are justified in taking the intuition as true. Then, I moved on to defend the emotion of agent-regret. There is a plausible argument that agent-regret is distinct from guilt, I held; and although the contention that agent-regret is irrational is somewhat convincing, it can be rebutted through appreciating the strategic rationality of the emotion in drawing the agent's attention to and motivating her to act upon the requirement to take responsibility for what an agent is merely causally responsible for.

I move on in the next chapter to consider some objections to my proposed solution.

Chapter Six: Objections to the NV View

Introduction

In the previous chapter, I expanded upon two concepts vital to Wolf's solution to the problem of resultant moral luck (RML), but of which she gave insufficiently detailed explications: agent-regret and the nameless virtue. I argued that we were justified on both intuitive and consequentialist grounds in taking there to be a moral requirement to take responsibility for what we are only causally responsible for; and that there are good reasons to consider the distress that arises in situations of unavoidable moral wrongdoing (UMW) and RML as distinct from guilt. I therefore concluded that it was plausible that Wolf's view could provide a way of reconciling the apparently conflicting intuitions RCW and MDJ.[128] In this chapter, I will address several possible objections that could be raised against the Nameless Virtue View (NV View.) Some of these objections are aimed at Wolf's argument generally, but are such that, if correct, would render my solution problematic. Others are at my application of Wolf's ideas to situations of apparent UMW.

The plan of the chapter is as follows. First, I address the Generosity Objection. This objection states that there is a contradiction in Wolf's account of the nameless virtue and the moral requirement it grounds (RTR.) Specifically, Wolf holds that it is a virtue due in part to the fact that it involves going beyond what is required of one (Wolf 2001, p. 14). However, if it involves going beyond what is required of one, it seems incoherent to hold that one can be morally required to act in accordance with it. I argue that this objection rests on a mistaken conception of the nature of the requirement.

Second, I tackle the Observer Puzzle Objection. This objection relates to the fact that we want an agent such as $Driver_1$ to feel guilt but also attempt to counsel him to reduce his level of guilt. I argue that this objection dissolves once one appreciates the way in which emotions become mixed up in tumultuous situations.

Third, I examine the Motivation Objection. The Motivation Objection holds that it is unclear how a strategically rational emotion could motivate an agent to take action. I argue that it fails to

---

[128] Recall that these intuitions state the following:

> An action that is all-things-considered morally right cannot also be morally wrong (RCW.)
>
> In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision (MDJ.)

appreciate that strategically rational emotions motivate action in a variety of contexts, and so it is unclear why it is problematic in this case.

Fourth, I grapple with the Objection from Tragic Cases. This objection concerns the fact that, in case of particularly tragic non-culpable harm, the agent experiences emotions so profound and deep that it seems implausible that they could be merely strategically rational. I contend that there are several convincing responses to this objection, and so it does not seriously imperil my view.

Fifth, I consider the Choice Objection. This objection states that if an emotion is strategically rational it is subject to the requirements of practical reason. But because experiencing agent-regret is not a choice (at least in the vast majority of cases) it cannot be subject to these requirements. I show that this argument ultimately does not go through.

Sixth, I examine the Disanalogy Objection. On this objection, RML and unavoidable moral wrongdoing are disanalogous due to differences in the position of the luck. In situations of RML, the agent's action precedes the luck; but in situations of apparent unavoidable moral wrongdoing, the luck precedes the action. This, it might be argued, renders Wolf's solution inappropriate to these situations. I argue that this is unjustified.

Finally, I consider the objection that Wolf's solution to the problem of RML merely resituates the luck inherent in morality. Because the agent does not choose to be subject to a moral requirement to act in accordance with the nameless virtue, this objection goes, Wolf's solution solves the problem of RML at the expense of allowing for circumstantial moral luck. This is an even bigger problem for my application of her solution to the problem of unavoidable moral wrongdoing, because unavoidable moral wrongdoing is partially problematic due to the way in which it allows for CML. I acknowledge that this objection is compelling, and set it aside for the following chapter.

6.1: The Generosity Objection

One objection that could be mounted against Wolf's account — and therefore to my argument against UMW — is the way it conceptualises the nameless virtue. Specifically, it seems problematic for her argument that she terms the nameless virtue to be a type of, or "akin to" generosity. Let us call this the Generosity Objection.

The Generosity Objection goes as follows. Wolf appears to see acting on the nameless virtue to be a moral requirement — she notes that we expect an agent such as Driver_1 to take responsibility for the consequences of his actions (in other words, to act in accordance with the nameless virtue):

> We expect the vase-breaker, like the truck driver who hit the child, to acknowledge that the consequences of his behavior have something specifically to do with him. We expect the vase-breaker to offer to pay, then, not only because we want him to be generous, but because we expect him to accept contingency in the determination and assessment of who he is.
>
> (Wolf 2001, p. 14)

In this passage (and throughout her essay), Wolf notes that we "expect" the agent to act in accordance with the nameless virtue. As "expect" here is clearly used in a normative sense and not a descriptive sense (as in, for instance, the sentence "I expect it will rain tomorrow"), and as a normative expectation seems synonymous with a moral requirement, it must be the case that Wolf holds that acting in accordance with the nameless virtue is a moral requirement. Therefore, it seems as if Wolf holds that generous action can be a moral requirement.

But this is problematic. Although the nature of generosity is a largely neglected topic in recent philosophical literature, it seems to be a conceptual impossibility for a generous action to be a requirement. Generosity appears supererogatory rather than obligatory.[129] Here is an example to illustrate this. When Peter Singer won the Berggruen Prize, a US$1 million dollar cash prize awarded annually to an intellectual whose ideas have had a meaningful impact on the world, he decided to give the entire prize away. Although Singer himself would (probably) disagree,[130] it seems odd to consider this the mere fulfilment of a requirement. If he had not given it away, for instance, few would have deemed him to have committed a moral wrong (although some might have condemned him for hypocrisy.) Similarly, when parents buy a car for their 17 year-old daughter, we comment that it was very generous of them. We do not mean by this, however, that buying the car was a moral requirement. If they hadn't bought the car, we would not consider them to have done anything wrong. Contrast this with this set of parents purchasing for their daughter a bag and stationery for school. We do consider the purchasing of such items to be a moral requirement, but we would not term it generosity. This suggests that generosity is not required. If the nameless virtue is a type of generosity it cannot be morally required for an agent to act in accordance with it. But this is problematic, for in situations of apparent UMW we do not consider the distress and subsequent moves to take responsibility for the consequences of one's actions to be supererogatory. We regard them as obligatory. Therefore, it appears that Wolf's account has a problem. Specifically, given what Wolf says

[129] See Stout (2015) and Miller (2018), both of whom hold that generosity is supererogatory. Stout appears to endorse Wolf's arguments about the nameless virtue, but does not address the discrepancy between his view that generosity is supererogatory and Wolf's view that it is, at least in cases of the nameless virtue, a requirement (see pp. 153-155.)

[130] See Singer (1973), for instance.

about the nature of the nameless virtue, it is hard to see how it can properly explain the phenomenology of situations of either UMW or RML.

One way of resisting this is to argue that it is not the generosity aspect of the nameless virtue that grounds the requirement to act in accordance with it, but the second aspect. The short passage quoted above suggests that, for Wolf, the agent ought to act in accordance with the nameless virtue not because of its relationship to generosity (we only "want" the agent to be generous), but because of what Wolf calls its "other aspect" — its benefits to psychic health. We "expect" — or require — the vase-breaker to offer to pay in order to acknowledge that his identity has been shaped by what he did (even though he is not morally responsible for it.) Therefore, perhaps it is not the generosity component of the nameless virtue that makes embracing it a requirement, but rather this second component. This, then, could be a way of escaping the Generosity Objection.

But such an argument misses the fact that this second component is simply a product of the generous action. That is, on Wolf's view the salubrious effect on one's "psychic health" is not some separate component which can be achieved independent of the generous action. The performance of a generous action is a necessary condition for this second, required, aspect of the virtue. Therefore, the problem remains.

Another — more promising — way to avoid this objection is to argue that some types of generosity are *not* supererogatory. In other words, there are types of what I think can reasonably be called generosity that *are* morally required. The story below illustrates this:

> The bus Bob is on is very full, and there are no spare seats available. Onto the bus steps Julie, a woman who is heavily pregnant and carrying a couple of bags of heavy groceries. All others on the bus are either elderly or in some physical state such that giving up their seats would cause hardship, whereas Bob is young and healthy. Upon seeing Julie, Bob offers her his seat, which she gratefully accepts.

It is plausible that Bob's act is an act of generosity. But it also seems intuitively correct that his act is morally required. To illustrate this, imagine that instead of offering Julie his seat, Bob had thought to himself, "I can see that this woman needs the seat more than I do, but to give it to her would be an act of generosity, and so isn't morally required. Because it's not morally required, I'm not doing anything wrong in not giving it to her. Therefore, I'm not going to give it to her." In this case, we would object to Bob's reasoning, and counter that he was wrong not to give up his seat, and that the generous action was required. Therefore, it seems that at least some types of generous action are morally required. If

this is correct, then it is the case that in some situations generosity is morally obligated, not merely supererogatory. Therefore, the "generosity objection" against Wolf's argument does not go through.

An alternative way of reconciling the apparently supererogatory nature of acts such as generous actions with the intuition that it can be the case that we ought to do them has been put forth by Julia Driver. On Driver's view, there is a category of actions that can be termed "suberogatory." These are actions that are morally bad, but which are nevertheless permissible. Her illustrative example is quite similar to the one I have used above. It involves a "train [that] is almost full, and a couple [who] wish to sit together, and there is only one place where there are two seats together… [but] the person ahead of them takes one of those seats." On Driver's view, "the [single] person on the train who refuses to take a less convenient seat — and thus greatly inconveniences others — has done something bad for which he can be blamed." (1991, p. 287) But he has done nothing wrong, because his action, although bad, was not impermissible. Driver calls such a situation a "morally charged situation," because it is a situation in which one can either act in a supererogatory or a suberogatory manner. The supererogatory action is morally good, and the suberogatory action morally bad; and so it is morally better for the agent to do the supererogatory action (*x*) and not to do the suberogatory one (*y*). In other words, says Driver, the agent ought to do *x* and ought not to do *y*. But because neither of the actions in a morally charged situation is impermissible the agent cannot do wrong in such a situation. Thus, the agent who takes a seat which deprives the couple of the opportunity to sit together does something morally bad, but not morally wrong. This, Driver claims, can resolve the tension between our sense that an agent is deserving of moral censure for not doing an act, and our sense that the doing of the act is optional. Applying this to the situation above, one might argue that Bob is in a morally charged situation, and so while it is not impermissible for him to refuse to give up his seat, it is nevertheless morally bad.

However, if Driver's view is correct, then it renders my definition of wrongdoing problematic. This is because it divorces "ought" (in its moral sense) from wrongdoing. In reference to the man on the train who selects his seat inconsiderately, depriving the couple of the opportunity to sit next to one another, Driver writes that the agent "has done something bad for which he can be blamed. But he has done nothing wrong." (1991, p. 291) The view of wrongdoing that I provided earlier (see 1.2), however, is as follows:

Wrongdoing occurs when an agent acts in a way in which she ought not to act.

Therefore, if Driver's account is correct, then my definition of wrongdoing is unacceptable and would need to be revised.

But I do not see any reason to accept Driver's account over my own. Driver's account requires one to accept that it can be the case that an agent can fail to act in accordance with what he has the most (moral) reason to do, and yet not do wrong. This is intuitively highly implausible. My suggestion for how to reconcile the apparently supererogatory nature of generous action with the fact that it is (in certain circumstances) required is both simpler and more convincing. On the view I laid out above, an agent such as Bob can face a situation in which he determines that the moral good of generous action (such as giving up his seat on the bus) outweighs other moral considerations (such as his right to occupy a seat on the bus), and so it is wrong if he does not perform the generous action.

Even if we accept that my view is preferable to Driver's, one could simply say that in situations like the one described above, Bob's actions are in fact not obligatory. Rowland Stout's interesting account of generosity would seem to lead to this conclusion. On Stout's view, generosity involves adopting in appropriate circumstances a role or status that one is not obligated to adopt.[131]He gives the example of a couple who allow their child to walk to school every day (it is through a safe neighbourhood and the route is short and so on.) One day it is raining, and a neighbour who is driving past in her car sees the child getting wet as she makes her way home. She picks the child up and gives her a lift. This is generosity because the neighbour is adopting the role of the parent in appropriate circumstances.[132] This seems to parallel the situation I outlined above; on this view, Bob adopts the role of one with special obligations to Julie (her husband, for instance) in appropriate circumstances. He is not obliged to adopt the role and its attendant responsibilities — to do so is supererogatory — but that he does adopt it is morally good.

The view that generosity involves adopting the role typically filled by another in appropriate circumstances is plausible. However, it is unclear why the adoption of such a role cannot be obligatory. Consider a revised version of Stout's example. Imagine the woman in the car observes the child walking home in the rain and recognises that it is appropriate for her to take on the role of parent, but remembers that a daytime soap she is mildly interested in is beginning soon, and that if she gives the child a ride she will miss the opening few minutes. As a result of this, she drives on past the child. On Stout's view, the woman does nothing wrong here — picking up the child is

---

[131] The distinction between adopting a role and adopting a status is temporal — adopting a role is short term (perhaps for a single action or series of actions), whereas adopting a status involves a longer term commitment.
[132] On Stout's view, adopting a role is appropriate when the adoption picks out an Aristotelian mean between "meanness" and "presumptuousness." How one should determine where this mean lies is not spelt out, but it seems as if he believes it is apparent through intuition. (Stout 2015, p. 149)

supererogatory. This is highly implausible. The minor benefit the woman gets from watching the first few minutes of a daytime soap does not seem to justify allowing the girl to walk home in the rain.

If this is correct, then Stout's account of generosity needs to be updated to reflect that sometimes one is morally required to be generous. Whenever generosity–the adoption of a role — is appropriate, it has positive moral value. This does not mean that one is required to adopt the role in all circumstances; perhaps there are other considerations which mean that, all-things-considered, one is not required to adopt the role. If, for instance, adopting the role of a parent and picking the child up would have caused the woman to be late for her own child's birthday party, picking the child up would not have been a requirement. But absent other morally salient considerations, the generous action *is* required.

There is however another counter-argument available to opponents of Wolf's view. It goes like this. Wolf characterises the nameless virtue as a virtue *because* it involves going beyond what is morally required of one. That is, its goodness lies in the fact that one is giving more than one is required to. However, if this is the case then it cannot be required — it seems to be a conceptual impossibility for one to be required to do more than one is required to do. This objection could also be levelled at my argument above: at least part of the moral goodness of Bob's giving up his seat for Julie lies in the fact that it goes beyond what is required of him. However, if that is the case then my view that it is a moral requirement cannot be correct. This gives credence to claims that Wolf's solution to the problem of RML — and therefore the use to which I put her solution — is problematic.

But this objection can be countered by examining Wolf's definition of generosity more closely. As she defines it, "generosity generally involves a willingness to give more — more time, more money, more love, more lenience, more, in one way or another, of oneself than justice requires." (2001, p. 14) On this view, it is not that generosity involves going beyond what morality *on the whole* requires of one, but rather that it involves going beyond what *one aspect* of morality —that is, justice — requires of one. Wolf's formulation is a highly plausible analysis of the concept of generosity. This formulation accommodates the intuition that generous action involves going beyond (certain) requirements, but also accommodates the (apparently conflicting) intuition that in some circumstances it can be all-things-considered morally required. Because justice is simply one consideration in determining the morally correct course of action, it can sometimes be morally right, all-things-considered, to act in ways that do not accord with it. To see this, consider again the revised version of Stout's scenario I outlined above. It is arguably the case that it goes beyond the demands of justice to require the woman driving to have to give someone else's child a ride home. That responsibility properly rests with the

child's parents. However, the demands of generosity make the act morally required.[133] The demands of justice, then, do not always accord with what is all-things-considered morally right. If this is correct, then the fact that acting in accordance with generosity is to go beyond what justice requires does not mean that it cannot be, all-things-considered, a moral requirement.

In this section, I have acknowledged that there is an apparent contradiction between the nature of generous action and the requirement for an agent to perform such an action. I have argued that this contradiction dissolves once one recognises that generous action does not consist in going beyond what one is morally required to do, but rather in going beyond what one aspect of morality — justice — requires of one. I conclude, then, that the fact that the nameless virtue picks out a moral requirement is not inconsistent with the fact that it is a type of, or closely related to, generosity. The generosity objection to my argument therefore does not go through.

6.2: Observer Puzzle Objection

Another objection that might be mounted against Wolf's account is a version of what has been called the Observer Puzzle. In a discussion of third-party observers to incidents such as that involving Williams's truck driver, Kamtekar and Nichols state the Observer Puzzle as follows: "why do observers judge both that agents should not feel guilty and that if they do not feel guilty, they are deficient in some way?" (2019, p. 183) It appears that, for a solution to the problem of RML to succeed, it must provide a way of resolving this apparent contradiction, and it is unclear how Wolf's solution does so.[134] In this section, I explain why this objection does not go through.

Domsky (2004) makes a version of the Observer Puzzle Objection with specific reference to Wolf's argument. He holds that the idea that "unlucky agents are no more blameworthy than their

[133] I am not assuming the truth of utilitarianism here. Rather, I am inclined to agree with Gowans that there is no systematic way of determining which moral principles to act on in any given situation, and that the correct method is a sort of Rossian consideration of the strengths of the relevant intuitions in any given situation (see Gowans 1994 p. 132)

[134] Kamtekar and Nichols's way of resolving the apparent contradiction is to distinguish between the actual domain (AD) of an emotion and the proper domain (PD). While the proper domain of fear is things (animals, situations, objects, people etc.) that are dangerous, the actual domain includes things which merely give the appearance of danger, such as a daddy longlegs spider or a ride to a great height in a glass elevator. Similarly (on the view of Kamtekar and Nichols), the proper domain of guilt is negative outcomes for which one is morally responsible, but the actual domain is outcomes for which one is causally responsible (that is, outcomes which one had control over but which one could not reasonably be held responsible for not foreseeing.) This explanation is unconvincing however. In the case of fear at a daddy longlegs, while we might consider it understandable, we do not consider it appropriate. But in the case of an agent such as William's truck driver, we do consider the emotion–moral distress–appropriate. Therefore, the proposed solution of Kamtekar and Nichols appears problematic.

lucky counterparts but that they still ought to blame themselves more than those others blame themselves" is "bizarre." (2004, p. 452) In addition, he argues that there can be no reason for such agents to begin their self-blaming if they are not actually more blameworthy, rendering the view even more problematic.

Such a criticism appears to have merit. Recall that, on Wolf's view, despite the fact that $Driver_1$ and $Driver_2$ are equally morally guilty, the one who ran over a child appropriately feels much worse than the one who did not. But Wolf's comments on how third-party observers ought to act towards them seem to contradict this. For instance, Wolf writes that, in the case of $Driver_1$, outside observers ought to counsel him, using arguments from the rationalist position,[135] against "judging himself too harshly," to show him "how limited was his faultiness and how large the element of luck" (2001, p. 10). However, she also writes that "we want the driver [who has killed a child] to feel worse." (2001 p. 10) Therefore, her account seems vulnerable to the objection that she provides no solution to the Observer Puzzle. If we want the driver to feel worse, why bother using rational arguments to bring down his level of guilt?

I think we can dissolve this puzzle if we view it in light of what Wolf later says about agent-regret and guilt. Wolf holds that

> when a faulty action causes harm, our feelings tend to get all mixed together — several sources for feeling bad about oneself and for being motivated to engage in certain sorts of actions combine to produce a mental state that is not itself analyzable into component states.
>
> (2001, p. 17)

In other words, an agent is not in a position to determine how guilty or blameworthy he is, because agent-regret and guilt blend together to form a practically indistinguishable melange. Thus, a third party "wants" an agent like $Driver_1$ to feel worse initially because this shows that he is feeling an appropriate total amount of agent-regret and guilt (although the proportion of each is not precisely determinable.) But it is also right for the third party to attempt to use rational arguments as mentioned above to help $Driver_1$ sort through his mixed emotions and therefore to arrive at a more appropriate approximation of his guilt. It is not therefore the case that Wolf's account cannot explain the Observer Puzzle. Contrary to Domsky's claims, the account provides a clear way of reconciling the apparently conflicting judgements, and also an explanation for why an agent would begin her self-blaming. Indeed, it is a point in its favour that it does provide a solution to the problem.

[135] Recall that the rationalist position is the view "according to which equal recklessness deserves equal blame." (Wolf 2001, p. 6)

However, it could be objected that while Wolf's account solves the Observer Puzzle for resultant moral luck, it cannot do so for cases of unavoidable moral wrongdoing. This objection goes as follows. In a case of resultant moral luck, the agent feels agent-regret and also (because he has committed, to use Wolf's phrase, a "faulty action") guilt, and these emotions get mixed together. It is therefore appropriate for observers both to hold that negative emotions are appropriate, and to counsel the agent in order to reduce his guilt to a level appropriate to his degree of wrongdoing. In cases of unavoidable moral wrongdoing, however, the agent (on my view) has not committed a faulty action, and so it seems as if his emotions of moral distress would be all agent-regret rather than a mixture of agent-regret and guilt.[136] It is unclear why the agent would be mixed up in his emotions, and therefore why it would be necessary for observers to counsel him. It should be clear to him that he has not committed wrongdoing. Observers should not see the need to counsel the agent to reduce the level of her feelings of guilt; and if they did, the agent should view such counselling as extraneous or even absurd. But this is not the case. In such situations, observers *do* feel the need to engage in such counselling, and it is hard to imagine the agent being puzzled by its purpose.

But this argument is unconvincing. On my view, while the agent has not committed wrongdoing, he experiences self-directed moral distress which he considers justified. Because the agent is performing an action which would normally be wrong, and because he is experiencing self-directed distress which he intuits is justified, it is natural for him to consider that he might have committed wrongdoing, even though this would conflict with the intuition RCW. It is for this reason that observers consider it appropriate both for the agent to require counselling, and for them to provide the counselling. An agent who required no such counselling would likely be experiencing no self-directed distress, and would therefore be responding inappropriately to the negative consequences for which she is causally responsible.

One final objection might relate to sophisticated moral agents. It goes like this. Imagine that the truck driver (let us call him Driver_p) is a moral philosopher, and has spent extensive time pondering issues relating to moral luck, moral responsibility, moral emotions and so on. Upon hitting the child, Driver_p experiences moral distress, but because he has pondered such matters at length, he is aware that what he is experiencing is agent-regret and not guilt, and so does not require any counselling in order to disambiguate the two emotions. In the aftermath of the event, when an observer attempts to engage in such counselling, Driver_p responds, "I only feel a small amount of guilt — my wrongdoing was relatively slight. You needn't counsel me on that point. It's tragic that the child

[136] The same is true of Williams's scenario involving the truck driver, of course, as in his original formulation the driver commits no negligence.

died, though, and I feel terrible agent-regret about being causally responsible for it." Were a person to give this response, we would likely find it inappropriate. This seems to show that it is not the inevitability of mistakenly feeling guilt that grounds our expectation that we will need to counsel agents such as $Driver_1$. We expect them to experience guilt regardless. If this is right, then this solution to the Observer Puzzler cannot be correct.

But this argument isn't convincing. If one were to encounter an agent such as $Driver_p$ and attempt to provide him with counselling only to find that he did not need it, I do not think we would find it problematic as long as we were sure that what he was saying was in fact true, rather than a cover for his failure to appreciate the significance of his causal responsibility for the tragedy. Obviously such surety would be difficult to come by; but if we possessed such surety, it seems as if we would find an agent's denial of guilt to be unproblematic.

Therefore, Wolf's solution of the Observer Puzzle succeeds in the context of both resultant moral luck and unavoidable moral wrongdoing, and so the objection appears unfounded.

6.3: The Motivation Objection

Another possible objection relates to my solution to the irrationality problem. Recall that, in response to the objection that agent-regret must be irrational, l argued that it was rational on strategic grounds (see 5.2.5). This solution could give rise to objections relating to agent-regret's ability to motivate the agent to take action.

As I argued in the previous chapter, agent-regret is strategically rational because it draws attention to and motivates the agent to act upon the requirement to take responsibility (RTR.) Its strategic rationality is similar to the strategic rationality of emotional responses to fiction, with these responses being rational because they enrich our lives. However, as I noted, when one responds emotionally to fiction, one is not motivated to act in a way that is at all similar to how one would be motivated to act if one were to experience the same emotion in response to a real world object. To take a famous example from the literature, if one sees green slime oozing towards her (or apparently towards her) on the screen at a cinema, she does not run from the theatre or alert the authorities.[137] If, then, agent-regret is merely strategically rational, it too should be expected to motivate minimal action on the part of the agent. However, my argument for its strategic rationality is that it motivates more than minimal action: it motivates acting upon RTR, which is sometimes quite demanding. As it

[137] The example is originally Kendall Walton's. (1978, p. 5)

is unclear that strategic rationality can motivate in this way, it seems as if my argument for the rationality of agent-regret founders.

My response to this objection is as follows. Note that in the case of emotional responses to fiction, it is not that the agent is motivated to take *no* action. To return to the example above, as the green slime oozes (apparently) towards her the viewer might cover her eyes, clasp the arm of her partner for protection, or let out a little gasp. She might even leave the theatre if her fear becomes too intense. Therefore, while the agent takes action, it is *different from*, although related to, the actions she would be motivated to take if the slime were really oozing towards her. The actions are milder than if this were the case. This suggests that, even when an emotion is strategically rational rather than representationally rational, it can motivate action. The same is true, I suggest, in the case of agent-regret. An agent who experiences distress at being causally responsible for an outcome is motivated to perform *some* actions that resemble in *some* ways the actions she would take if the cognitive component of the moral distress involved the judgement she had done wrong. However, what she is motivated to do is substantially milder than if the cognitive component accompanying the affect indicated wrongdoing. Let us return to the example of Craig and the two young women whose car he takes without permission. In this case, Craig experiences an affect of moral distress with a crude appraisal such as "I've done wrong," or "morally bad!" This appraisal is then refined through a cognitive process that leads him to recognise that while he has produced morally bad consequences, he has not done wrong . It is this evaluation — incorporating the thought that one has done something morally bad and also an awareness that this is not strictly true — that motivates him to take action subsequently, in particular to apologise and explain his actions to the young women and perhaps to buy them flowers or some other sort of gift. It does not motivate action in the same way as an affective appraisal of moral distress coupled with the cognitive component that one has done wrong — had Craig's action in taking the young women's car been wrong, all-things-considered, and had he cognitively evaluated it as such, he would have been motivated to act in ways beyond simply offering an explanation and some flowers. He would also have resolved to behave differently in the future, to turn himself in to authorities, and to accept punishment, for instance.[138] As with an emotion experienced in response to fiction, then, moral distress that is strategically rational — in other words, agent-regret — motivates a response that is *like* that of the representational rational emotion, but which is different (milder) in important respects.

[138] This assumes, of course, that Craig is properly sensitive to his moral reasons for action.

6.4: Objection from Tragic Cases

A related objection concerns the phenomenology of agents who are non-culpably responsible for especially tragic outcomes. One such agent is Sophie. This objection holds that the moral distress such agents experience — so agonising, so engulfing — is too profound and resistant to mitigation for it to be merely strategically rational. When we respond emotionally to a fictional character, the emotion may be quite intense; however, it is typically short lived, and it is possible to diminish the emotion by reminding ourselves that neither the character nor the situations she finds herself in are real (and that therefore her emotion is not representationally rational.) However, Sophie's distress is not short lived; and it seems implausible that she could quell her distress to any great extent by reminding herself that it isn't representationally rational. Therefore, one could conclude, Sophie's distress is not merely strategically rational.

The charge that the duration of Sophie's distress makes its strategic rationality problematic seems unwarranted. After all, while emotions experienced in response to works of fiction are often short lived, they are not always so. For instance, I read J. M. Coetzee's novel *Disgrace* many years ago, but I nevertheless still experience pity for the main character, David Lurie, when I contemplate the novel's final scenes.

But the objection that Sophie's distress is impervious to amelioration in the way that emotional responses to fiction are is more convincing. It is undeniable that Sophie's reminding herself that her negative emotion was, in a sense, inappropriate would not cause it to subside substantially. However, there are a few possible responses to this, which together seem to overcome the objection. First, it is widely recognised that acknowledging an emotion as irrational does not necessarily cause it to subside much. Take someone with a fear of heights. Were he to find himself on the observation deck of the Burj Khalifa in Dubai, 125 floors up, he would likely feel immense fear. This fear would likely persist to a large extent even if he were to remind himself of the fact that he was safely ensconced inside, behind thick glass windows, and unlikely to find himself suddenly outside, plummeting to the ground. Therefore, it is unclear that it is a problem for my view that Sophie's moral distress would not diminish substantially.

Second, recall that Sophie has had her child snatched away from her, presumably to be murdered. Therefore, it is to be expected that distress will persist, because distress at the fact that her child has been taken off to be killed is appropriate. It is the self-directed component of the distress whose rationality is strategic. The distress at the abduction and likely murder of her daughter is

representationally rational, and so recognition of the representational irrationality of the self-directed moral distress would leave this distress unaltered.

Third, it is unclear that it is correct that if Sophie were to remind herself of the representational irrationality of her moral distress it would not diminish it or change its character. While it would no doubt remain intense (for the reasons outlined above, as well as due to its strategic rationality), it seems to me that it would have some ameliorative effect. Imagine that Sophie is in the grip of intense moral distress, wondering what she has done and castigating herself for her actions. The thought then enters her mind that her regret is, in a sense, irrational — she made the best decision possible and so has nothing to regret. Is it really plausible that the awareness of this does *nothing* to mitigate her moral distress? I submit that it is not. Her moral distress will not dissipate; indeed, on the view I have argued for it would be irrational for her to try to eliminate it entirely. But it seems plausible that it does diminish, at least somewhat.

I therefore conclude that the Objection from Tragic Cases does not seriously imperil my view.

6.5: The Choice Objection

Richard Joyce (2000) argues that the strategic rationality of emotional responses to fiction lies in the fact that one can choose to experience them. For this reason, such responses are subject to the norms of practical rationality rather than doxastic (epistemic) rationality. His view is as follows. Because consuming a work of art is a choice, and because one knows that emotion will arise from consuming this work of art, then the emotion itself is a choice. Furthermore, Joyce argues that

> Much more often than not ... we are in firm control of our emotional states when embroiled in a fiction. Usually, the music makes me cry because I am *letting* it, the ghost story makes me fearful because I am *encouraging* it [emphasis in original.]
>
> (2000, p. 217)

We choose to allow ourselves to experience emotional responses to fiction in two senses, then; and these responses are therefore subject to assessment according to practical reason. But, it might be objected, the moral distress an agent such as Sophie experiences is not something one can choose to experience. Sophie does not choose to have a Nazi guard accost her and force her into an agonising and cruel decision. Therefore, the rationality of emotional responses cannot be used to explain the rationality of moral distress in situations such as that Sophie faces.

One way of responding to this objection would be to simply to hold that pragmatic or practical reasons encroach upon theoretical justification.[139] But there is another possible explanation for why emotional responses to fiction are strategically rational: It is strategically rational because, once the emotion has arisen, one has no reason to take steps to dismiss the emotion. Recall Joyce's principle 7:

> If a person performs action φ, believing that ψ will result — and no other agency is involved — then, if ψ does result, that person is deemed not merely responsible for ψ, but to have performed the action ψ.
>
> (2000, p. 218)

An agent who causes non-culpable harm does not (definitionally) choose to cause the harm. Nevertheless, once she has caused the harm and is in the grips of distress, she does not take steps to dismiss the emotion (by telling herself that it is irrational, by not thinking about the incident, etc.) Because it is controlled (to some extent, at least), it is therefore appropriately subject to assessment according to practical reason after all.

6.6: The Disanalogy Objection

A further objection that could be made is that cases of RML and cases of UMW are disanalogous due to differences in the nature of the luck. Let us call this the Disanalogy Objection. In this section, I argue that this objection does not go through.

The Disanalogy Objection is as follows. In cases of RML, the action comes first and the luck comes later. However, in cases of UMW, the opposite is true.[140] This, it could be argued, makes the situations disanalogous in ways that render the application of Wolf's solution to situations of UMW problematic. In cases of RML, an agent makes a decision which seems to her rather mundane, but which is later rendered momentous and tragic through bad luck. Driver_1, for instance, does not consider his decision to delay having his brakes fixed to be a particularly significant one, as it is unlikely it will lead to any particularly terrible consequences. It is only the subsequent ingredient of bad luck that makes the decision seem awful. Cases of UMW, however, are the opposite of this, in that the bad luck comes first, and the agent's decision later.[141] Consider Sophie's plight. It is very bad luck that she should find herself in a Nazi concentration camp and encounter a sadistic guard who decides

---

[139] For more on pragmatic encroachment, see Fantl and McGrath (2002) and Fantl (2012.) Closely related is recent flurry of literature on moral encroachment. See, for instance, Basu (2018, 2019.)

[140] This is noted in Kamtekar and Nichols (2019.)

[141] In this way, situations of apparent UMW are actually cases of circumstantial moral luck. Nagel notes as much in his essay on moral luck. (1979, p. 212 footnote)

to make her sacrifice one of her children. In this case the luck precedes the decision, and so when Sophie has to make her choice, she is in full possession of the facts and is explicitly and consciously aware of the consequences of her decision. This is not true of $Driver_1$, whose apparently only mildly negligent decision is subsequently twisted into a tragic and grave one whose consequences he never endorsed. To use Wolf's phrase, it is arguably within the sphere of Sophie's will for one of her children to perish, but not within the sphere of will of $Driver_1$ to kill a child. Sophie actually makes the choice to hand one of her children over to the Nazis to be exterminated. The truck driver does not choose to run over a child. For this reason, an opponent of my argument might claim, UMW and RML are disanalogous, and so my solution to the problem of UMW does not work. Agent-regret and the nameless virtue might explain the phenomenology of cases of apparent RML, but, it could be concluded, in cases of unavoidable moral wrongdoing the moral distress is appropriate because it picks out wrongdoing.

However, such an argument is unconvincing. On the view sketched out in the objection above, Sophie commits a greater wrong than $Driver_1$ because she willed the bad outcome — the death of her child — into existence. It is because of this worse will, then, that she is guilty of greater wrongdoing. But to claim that Sophie's will was worse than that of $Driver_1$ seems obviously false. The will of $Driver_1$ is to take the chance that his brakes will fail due to improper maintenance and cause grave consequences. The will of Sophie is to try to save one of her children instead of letting both face almost certain death. If one's quality of will is what matters, it seems as if Sophie's is superior to that of $Driver_1$. After all, Sophie makes the morally correct decision, whereas $Driver_1$ does not. Intuitively, it seems highly implausible that Sophie's degree of wrongdoing is greater than that of $Driver_1$. It is therefore unclear on what basis one could hold that Sophie had committed a greater wrong than $Driver_1$. But if this is correct, then it seems there is no basis to claim that the situations are disanalogous in such a way as to prevent the application of Wolf's ideas to situations of UMW.

One might respond that there is still a crucial difference in cases of apparent RML and cases of unavoidable moral wrongdoing: that the agent makes a choice to directly and knowingly produce the consequences in the latter, whereas in the former the agent only knows of the *possibility* of the consequences. Perhaps the wrongness of an act is linked not to the agent's intention but to what the agent expects (or should reasonably expect) to be the outcome of the act. If this is correct, then because Sophie knows that a morally bad outcome — the death of Eva, her daughter — will almost certainly result from her actions, she does greater wrong than $Driver_1$.

But it seems implausible that the degree of wrongdoing could be linked to the likelihood of the action producing a morally bad outcome. Consider the following story:

> Jones is an assassin who is engaged by a criminal gang to shoot a police chief who is making a concerted effort to stamp out organised crime. Jones has developed a bad tremor in his hands, and because of this, he knows that his shot is unlikely to be particularly accurate, and that there is a good chance his shot will miss entirely (assume that there is no chance of any bystanders being hit by the bullet or harmed in any way by Jones's inaccuracy.) Despite his ailment he is determined to do his best to kill the police chief. However, as expected, his shot misses, and the police chief escapes unscathed.

On the view articulated above, because Jones has developed a tremor in his hands which he knows will render his shot less accurate, he does less wrong in trying to kill the police chief than if he had not developed the tremor. But this seems highly implausible. Intuitively, Jones's act is wrong because he intends to kill a person. The fact that he knows the consequences of his actions are unlikely to be deadly seems to do little, if anything, to ameliorate the wrongness of the act. If this is correct, then it seems as if it cannot be the case that an act's wrongness accords with what the agent reasonably expects, or should reasonably expect, the consequences to be. This appears to render this version of the Disanalogy Objection unconvincing.

There is a further response to the Disanalogy Objection. On the objection, luck subsequent to an action can appropriately engender agent-regret, but not luck prior to an action (because in this case the agent has done wrong, and so the moral distress he experiences is guilt.) But if we consider a slightly revised version of Sophie's predicament, this appears problematic (let us call the Sophie of the revised situation Sophie_2):

> Sophie_2 is in a Nazi concentration camp and is accosted by a sadistic guard. The guard says that he will murder one of her children, and demands she choose which of her children will survive. Sophie agonises before choosing, at random, her son, Jan, as the one who should survive. However, the guard mishears Sophie and thinks she chooses her daughter, Eva. He therefore rips Jan away from her, whisks him off and murders him.

In this story, Sophie_2 does not intend to sacrifice Jan. It is luck subsequent to her action which produces this outcome. If, then, luck subsequent to an action can justify agent-regret but not guilt, the phenomenology of this situation should be different from that of the original situation. In particular, it seems as if Sophie_2 should feel additional distress — distress at the wrongdoing involved in intending for her daughter to die, and another lot of moral distress for being causally responsible for her son's murder. But this seems implausible. If we compare Sophie (the version from Styron's novel) with Sophie_2, we do not have the intuition that it would be appropriate for Sophie_2 to feel

additional distress. It seems as if it would be appropriate for her to feel the same level of distress as Sophie. But if this is the case, then the nature of the luck cannot be claimed to produce meaningful disanalogies. If this is correct, then RML and UMW should not be claimed to be disanalogous on that basis.

It seems as if there are no good reasons to accept the disanalogy objection, then. I will now move on to consider rejections relating to the concept of agent-regret.

6.7: Objection from the Resituation of Luck

Another possible objection has to do with the way in which Wolf's solution–and my application of it — seem to merely resituate luck, rather than eradicating it from moral evaluation. In other words, one sort of moral luck is apparently replaced with another. Moore has made such an argument. He writes as follows:

> [T]hose in Wolf's position—the anti-moral-luck camp—are primarily … motivated by the thought that it would be unfair to blame people even in part based on factors (like results) that they do not fully control. Yet notice this worry evaporates for Wolf's nameless virtue: we ought to feel agent-regret whenever we have non-culpably caused harm, despite the 'contingency' involved in such causation.
>
> (Moore, 2009, p. 33)

In other words, on Wolf's account a person is required to take responsibility for the consequences of their actions, even when she is not morally responsible for these consequences. However, the person has not chosen to be in a position in which she is required to take this responsibility. It is therefore merely (bad) luck that the agent should be subject to this moral requirement.[142] At least some agents who are subject to these moral requirements will fail to fulfil them, and therefore render themselves more blameworthy than agents who would also have failed to fulfil them, but did not commit non-culpable harm and so did not face such requirements. Hence, Wolf's account appears to deny that agents can be subject to RML but at the expense of exposing them to circumstantial moral luck (CML). However, if CML exists, then perhaps there are no good reasons for dismissing the possibility of RML. If this is correct, then my argument seems to falter.

---

[142] This type of moral luck–moral luck in regard to the obligations an agent faces–Zimmerman calls deontic moral luck (2019, p. 219-220). He distinguishes this from what he calls hypological moral luck — luck regarding what one is responsible for. Of course, given that an agent facing a more onerous set of obligations is more likely to do wrong and therefore to be morally responsible for a bad outcome, these two types of moral luck are closely connected.

The objection is perhaps even stronger when levelled against my application of Wolf's ideas to the problem of unavoidable moral wrongdoing. This is because if UMW is possible then situations of UMW are also cases of CML.[143] Although the fact that this is the case is not one of the primary reasons I gave in Chapter One for finding UMW problematic, the existence of CML would undermine the control principle (CP); and the plausibility of CP *is* one of the reasons I provided earlier in support of the plausibility of the intuition RCW. Therefore, if the solution I propose to the problem of unavoidable moral wrongdoing is reliant on the existence of CML, it undermines one of the reasons for seeking a solution to the problem in the first place. A critic might wonder why I did not simply dismiss the intuition RCW as insufficiently well-grounded and accept that UMW was possible.

There are, then, two forms of this objection — one which is levelled against Wolf's account, and one which is levelled against my application of Wolf's account. One way of refuting both forms is to argue that, despite appearances, Wolf's account does not require that CML be possible. But it is not immediately clear how one could take this line convincingly.

Can one find more success in addressing the two forms of the objection individually? One way of defusing the first form — that Wolf's solution is self-defeating because it solves the problem of RML by requiring CML — is simply to bite the bullet and accept that moral evaluation is subject to CML but not RML. After all, Wolf states explicitly that her account aims to solve the problem of resultant moral luck. She makes no claims to solve the problem of circumstantial moral luck. There seems nothing prima facie inconsistent with this position. But while there is no prima facie inconsistency, some philosophers have argued that this view — dubbed the "asymmetry view" — is incoherent.[144] If this is the case, then Wolf's solution is problematic after all.

Nor is there an obviously convincing response to the second form of the objection, the one relating to my application of Wolf's solution to the problem of unavoidable moral wrongdoing. Several possibilities present themselves, but none is obviously correct. First, perhaps CP is somehow coherent with the possibility of CML. A second possibility is that CP is false, but RCW retains enough plausibility from other sources that we are nevertheless justified in dismissing the possibility of unavoidable moral wrongdoing. However, because both of these possibilities seem on first inspection to be somewhat unlikely, if one wishes to adopt one of these positions the onus is on her to provide an argument showing its veracity. Because I have not made such an argument, an opponent is justified

[143] At least, this is the case with regard to situations of unavoidable moral wrongoing in which the agent is not responsible for the fact that he is in the situation. Situations such as Hursthouse's scenario involving a man who has promised to marry two women, but who can only marry one, are not cases of CML. (1998)
[144] See Hartman (2017, 2019a.)

(at this point) in viewing my solution to the problem of unavoidable moral wrongdoing with some scepticism.

To sum up, it appears that allowing for the possibility of CML, then, is incompatible with my solution to the problem of UMW. But it also seems as if my solution requires it. So, in order to salvage my solution, I need to do one of two things. First, I could argue that, contrary to appearances, the requirement to act in accordance with the nameless virtue is not an example of CML. Second, I could argue that CML differs from RML and is therefore unproblematic (or at least, that some types of CML are unproblematic, or that it is unproblematic in particular circumstances.) Addressing this issue is the subject of the next chapter.

Conclusion

In this chapter, I have argued against a number of possible objections, both against Wolf's account generally and against my application of it to situations of unavoidable moral wrongdoing. I contended that all but one of these objections did not stand up to scrutiny.

I ended the chapter by addressing an objection related to the resituation of luck that my solution seems to involve, acknowledging that this appeared problematic. The problem lies in the fact that, if my solution merely involves replacing one sort of moral luck with another, the solution would seem to violate the Control Principle. As the Control Principle is one of my reasons for considering the intuition RCW to be plausible, the argument seems self-defeating. Therefore, if my solution is to prove plausible, I need to show how at least some forms of circumstantial moral luck do violate CP. I turn to this task in the next chapter.

Chapter Seven: The Problem of Circumstantial Moral Luck

Introduction

In Chapter Four, I argued for a solution to the problem of unavoidable moral wrongdoing that reconciled the intuitions RCW and MDJ.[145] Specifically, I applied to the problem of unavoidable moral wrongdoing Susan Wolf's solution to the problem of resultant moral luck, arguing that the moral distress agents feel in such situations is best understood as agent-regret, and that there is a moral requirement for agents to take responsibility for what they are merely causally responsible for.

In the previous chapter, I outlined several possible objections against my view, and argued that almost all of them were unconvincing. But I acknowledged the plausibility of one of these objections: that my solution appeared to require that circumstantial moral luck be possible. This is because an agent who is causally responsible for bad consequences is, on my view, subject to moral requirements to which an agent who is *not* causally responsible for any bad consequences is not subject. An agent who does not fulfil these requirements is blameworthy; and so it seems as if Wolf's solution removes resultant moral luck from the picture, but at the expense of exacerbating the role of another form of luck–circumstantial luck–in shaping moral appraisals. Of course, circumstances outside the agent's control dictate her moral obligations in a whole range of contexts; Wolf's solution to the problem of resultant moral luck simply adds one more way in which this is true.[146] However, this simply raises the question as to why, if luck is so pervasive in establishing our obligations — and therefore seemingly in shaping our praise- and blameworthiness — resultant moral luck should be considered so problematic. This line of thought suggests that we should simply accept that luck plays a role in moral evaluation. If this is correct, then Wolf's solution to the problem of resultant moral

---

[145] Recall that these intuitions state the following:

RCW: An action that is all-things-considered morally right cannot also be morally wrong
MDJ: In (some) situations in which an agent has two conflicting obligations, it is appropriate for the agent to experience negative self-directed moral emotions even if she makes the all-things-considered correct moral decision

[146] David Enoch notes as much: "What you are morally required to do depends on the circumstances you are in, and many of those are not under your control; they are for you a matter of luck. So it is important to distinguish between (purported) cases of moral luck on one side, and cases of plain luck that has moral implications, or morally significant plain luck on the other. And while the latter may also be problematic, it is not as problematic as the former, and anyway its uncontroversial existence does not entail the highly controversial existence of the former." (2012, p. 131) But Enoch's view that we can deny circumstantial *moral* luck but accept morally significant plain luck is problematic, as I argue below (see 7.1.1).

luck is unmotivated, and the application of her solution to the problem of unavoidable moral wrongdoing also seems shaky.

I will briefly elaborate on this. First, if Wolf's solution does not solve the problem it was intended to solve, it seems questionable whether it could solve a different problem. Second (and more troublingly), if we accept circumstantial moral luck then it is unclear why unavoidable moral wrongdoing is problematic. In Chapter One, I gave the Control Principle as one of the key intuitions cohering with, and therefore supporting, the intuition RCW. But if circumstantial moral luck is possible, then CP is false, and so this supporting intuition is withdrawn. Indeed, if we accept circumstantial moral luck this could support the contention that unavoidable moral wrongdoing is possible, because unavoidable moral wrongdoing would just be another form of circumstantial moral luck. Allowing for circumstantial moral luck therefore seems to demotivate the project of seeking to deny unavoidable moral wrongdoing, and to point in favour of simply accepting it.

The aim of this chapter is to address these concerns. There are two possible ways of doing this. First, one could argue that, while the specific moral requirements an agent faces are (at least to some extent) outside her control, this does not affect the way her actions are morally appraised relative to others. In other words, one could deny that circumstances outside the control of the agent can affect the agent's degree of moral responsibility, and therefore deny that there is extant circumstantial moral luck. Michael Zimmerman (1987, 2002) and David Enoch (2019) have made arguments along these lines. If this view is correct, then Wolf's solution to the problem of resultant moral luck is not undermined, because there is no circumstantial moral luck, either. Nor is the Control Principle imperilled, and so the intuition RCW retains its initial plausibility.

Second, one could concede that specific moral requirements that the agent faces for reasons outside her control *do* affect the way she is morally appraised, but that this is not inconsistent with the denial of resultant moral luck. On this view, resultant moral luck differs from circumstantial moral luck in important respects, and so circumstances can affect blameworthiness (and praiseworthiness), but outcomes cannot. This is known as the asymmetry view. One could further hold that allowing for circumstantial moral luck is consistent with a revised version of the Control Principle, and that this version supports the intuition RCW.

In this chapter, I argue that a version of the latter view is correct. There is some extant circumstantial moral luck, but the type of circumstantial moral luck that is possible is consistent with a version of the Control Principle that supports the intuition RCW.

The plan of the chapter is as follows. I begin by examining the first strategy I mentioned above: denying circumstantial moral luck. I consider arguments from David Enoch and Michael

Zimmerman that, despite appearances, there is no circumstantial moral luck. I argue that both are ultimately implausible.

Next, I move on to examine the asymmetry view. This is a widely held position amongst philosophers, and I argue that a recent attempt by Hartman to reject it fails. I also argue that Rivera-Lopez's argument that there is a relevant difference between circumstantial moral luck and resultant moral luck provides us with reason to accept the view. I also argue that the acceptance of circumstantial moral luck is consonant with a plausible version of the Control Principle, one that can provide support for the intuition RCW.

Finally, I consider a few objections to my argument.

7.1: Response 1: No Circumstantial Moral Luck

As I laid out in the introduction to this chapter, if there is circumstantial moral luck (CML) it problematises my solution to the problem of unavoidable moral wrongdoing (UMW.) One way of defusing this problem is therefore to argue that there are good reasons for thinking that CML is not possible. This line of argument holds that although it appears that CML is possible, that appearance is deceptive, and the degree to which an agent is blameworthy[147] cannot actually be affected by circumstances outside of his control. On this view, then, an agent who for reasons outside of his control is acting in accordance with the requirement to take responsibility (RTR) and who fails to do so is subject to the same moral evaluation as someone who in the same circumstances would also have failed to act in accordance with the nameless virtue.

In this section, I examine two accounts denying the possibility of CML. First, I analyse David Enoch's view that, while there is "morally significant plain luck," there probably aren't good grounds for holding that there can be moral luck. Then, I consider the counterfactual view espoused by Michael Zimmerman. I hold that neither view is persuasive. Therefore, I conclude, denying CML is not a promising strategy to salvage the plausibility of my solution to the problem of UMW.

[147] Throughout this chapter, I typically discuss moral luck in terms of an agent's blameworthiness. At times, however, I also discuss it in relation to an agent's degree of moral responsibility. Because I take an agent's degree of moral responsibility to simply refer to an agent's degree of blame- or praiseworthiness, nothing hangs on the term being used.

7.1.1: Morally Significant Plain Luck: Enoch's View

An intriguing recent argument for doubting the possibility of circumstantial moral luck comes from David Enoch (2019.)[148] Enoch argues that the moral luck debate has failed to take into account the distinction between moral luck and what he calls "morally significant plain luck." Enoch uses an analogy between rationality and morality in order to argue that, while there is morally significant plain luck, there are good reasons to doubt that moral luck is possible. In this section, I first briefly lay out Enoch's account. Then I argue that it fails to provide convincing reasons to doubt that circumstantial moral luck is possible, and so does not provide a satisfactory way of responding to the charge that my solution is problematic.

Enoch starts by noting that "circumstances that are not under the relevant agent's control, in any relevant sense of control [have] moral implications." (2019, p. 258) That is, much of the time the circumstances we face are not of our choosing; and as what we morally ought to do is contingent on the circumstances we are in, what we morally ought to do is in large part determined by luck. But this type of luck, according to Enoch, is not moral luck but rather "morally significant plain luck." Morally significant plain luck, on Enoch's view, is luck which is "external to morality." This is in contrast to moral luck, which if it exists is internal to morality. By this, Enoch seems to mean that morally significant plain luck shapes the requirements that morality imposes upon agents (and so is external), whereas moral luck is luck in the degree of blameworthiness, praiseworthiness, and moral responsibility an agent deserves given the set of moral requirements that exists. (Enoch 2019, p. 259)

Enoch then uses an analogy from rationality to make his argument about moral luck. Enoch argues that what he calls "normative rational assessment" — what rationality requires of one — is (at least partially) dependent upon luck, but that "critical rational assessment" (how well one fulfils the demands of rationality) is not. It is plausible, Enoch argues, that moral assessment exists because "[w]e need to know who can be trusted to play well across a wide variety of morally relevant hands." (2019, p. 264) That is, just as in card games involving a mixture of skill and luck we evaluate the player based on how skilfully he played his cards, so too in morality we evaluate the agent based on the degree to which she acted in accordance with the moral reasons she had.

Enoch's view, if correct, would provide a way of denying that my solution to the problem of UMW requires that circumstantial moral luck be possible. On this view, an agent who faces moral

[148] Enoch's account is intended to provide reasons for doubting that either circumstantial or resultant moral luck is possible (he is less certain about whether it provides reasons to doubt constitutive moral luck.) However, as my purpose here is to analyse whether we have good reasons for doubting that circumstantial moral luck is possible, I will focus on his argument as it relates to this.

obligations due to her being causally responsible for some bad outcome is subject to morally significant plain luck, in that luck determines the moral obligations she faces. However, how she is morally evaluated — her degree of moral responsibility — is immune from the effects of luck.

But Enoch's view is implausible. Enoch holds that his account shows that we have good reason to doubt that moral luck exists — our critical moral evaluations are evaluations of how well the agent acted *given the circumstances*. In other words, because the circumstances are factored into the evaluation, luck plays no role in the way in which one is morally evaluated. But it is hard to see how this is the case. A chess analogy allows us to perceive this. Consider two players, A and B. A is assigned a set of three matches against grandmaster Magnus Carlsen. Carlsen proceeds to destroy A in all three matches, with his subtle moves leading A into a series of blunders. When A's moves are subjected to analysis by a sophisticated chess engine — let us say AlphaZero — it shows that A played with only 45% accuracy.[149] Meanwhile, B is matched up in his series against Homer Simpson. Although aware of the rules of chess, Simpson has little grasp of even basic strategy and plays moves almost at random. As a result, B defeats him in short order in each of their three matches, receiving an accuracy score of 98% from AlphaZero in the process.

Let us now engage in a critical rational evaluation of the play of A and B. If we compare each player's moves to the optimal moves as identified by AlphaZero we see that B played with far greater accuracy than A. In fact, B played almost perfectly. But because the moves of Homer Simpson were so disastrous, it did not take particular brilliance or insight to make the optimal moves. A, on the other hand, faced arguably the greatest player in chess history, and so finding the optimal way of combating his tactics was much more difficult. Therefore, even though a critical rational evaluation of A and B would find that B followed the normative demands of rationality far better than A, it is hard to conclude that luck played no role in the production of these assessments. After all, if A could have played his series of matches against Homer Simpson, he may well have made the optimally rational move almost every time, as well.

The parallel to morality is clear. Some agents, due to luck, face moral situations as challenging as a chess matchup against Magnus Carlsen. Let us take the character of Hanna from Bernhard Schlink's novel *The Reader*. Due to the circumstances of her upbringing, she is illiterate. When she is offered the opportunity of promotion in her job as tram conductor, she quits because she fears that her illiteracy may be exposed, a prospect she finds highly mortifying. She then accepts a job as a guard

[149] Chess engines assign players accuracy scores by comparing the moves of a player with the moves the engine would have made in exactly the same circumstances. Therefore, if a player is assigned a score of 100%, then that means a player has made exactly the moves the engine would have made.

in a Nazi camp, helping (albeit as a mere cog) to implement the Nazis' evil schemes. It is I think uncontroversial to say that Hanna behaves wrongly and deserves a negative critical moral evaluation. Hanna did not have to take a job in a Nazi camp; and although her illiteracy might have made it more difficult for her to know the full extent of the evil she was a part of, it seems appropriate to say that her epistemic position was such that she knew or should have known that working for the Nazis was wrong. But now let us imagine that, rather than living in Nazi Germany, Hanna lived in modern day Switzerland. Had she become jobless due to illiteracy, she would have had options to remedy her illiteracy, and may well have negotiated the joblessness without any wrongdoing at all. Even if she did not, it is highly unlikely her behaviour would have deserved a critical moral evaluation as damning as that deserved by one who participates in the genocidal schemes of the Nazis. Therefore, even if Enoch is right that agents' blameworthiness is determined by how well they play their moral hand, it seems that luck plays a significant role in an agent's evaluation.

Here is a second situation to illustrate the point. Imagine two women, A and B, who go out of town on a business trip. Both go to the hotel bar one evening for a quiet drink before bed. Coincidentally, the high school boyfriend of A (whom A has never entirely forgotten) shows up at the bar, and through weakness of will A shares a brief kiss with her former boyfriend before overcoming her weakness and retiring alone to her room. B also has a high school boyfriend whom she has never entirely forgotten, but he does not show up at the bar that night. Therefore, B finishes her drink and goes to bed as planned. In this situation, B makes no moral mistakes; to use Enoch's metaphor, she plays her hand perfectly. However, this is simply because her metaphorical opponent is more like Homer Simpson than Magnus Carlsen. The morally significant circumstances she faces are particularly unchallenging to navigate. A does not act as an ideally moral agent would–that is, she does not play her moral hand flawlessly–but she faces a much more difficult set of morally relevant circumstances. If A's former boyfriend had not materialised in the lobby, she would have played a flawless moral hand just as B did.

Perhaps Enoch could counter that critical moral evaluations should be adjusted to account for the difficulty of the situation. On this view, one's evaluation would be adjusted to account for the difficulty of the situations one faced. If, for instance, one faced a very difficult morally significant situation and one played one's metaphorical hand reasonably well, the raw evaluation could be adjusted upwards to account for the difficulty of the situation. However, if this is Enoch's view, he owes us an explanation of exactly what formula we should use to determine these evaluations, and how it would eliminate entirely the role of luck in appraisals of moral responsibility.

It seems, then, that if we rely on critical moral evaluations of agents' actions, there is still moral luck. That is, luck still plays a role in how one's actions are morally evaluated, and not simply in establishing what one's moral requirements are. It appears that what Enoch needs to argue to rescue his theory is that if agent A plays his moral hand as well as agent B *would* have played the same hand, then A and B are blameworthy to the same degree. In other words, his view appears to devolve into the counterfactual view.

Whether or not Enoch's view proves useful in solving the problem of circumstantial moral luck, then, depends upon whether the counterfactual view can be defended. I turn to this view in the following section.

7.1.2: The Counterfactual View

In this section, I will consider a way of denying circumstantial moral luck advanced by Michael Zimmerman (1987, 2002, 2015, 2019).[150] This argument for denying CML is known as the counterfactual view. Zimmerman marshals this view against what he calls "situational moral luck," a phrase he uses to encompass both constitutive and circumstantial luck. Because my purpose is not to consider whether constitutive moral luck is possible, I will restrict my analysis of the argument to its effectiveness as a strategy for denying the possibility of circumstantial moral luck. In the first part of this section, I lay out Zimmerman's counterfactual view. After that, I argue that it fails as a strategy for denying circumstantial moral luck it implies that all (or most) agents are infinitely blameworthy; and it saps blame of its communicative function.

The counterfactual view denying CML begins with CP and what Zimmerman holds is a related principle, which he terms the Larger Truth (LT). He expresses LT as follows:

> **LT:** The degree to which one bears culpability cannot be affected by something that neither is nor was in one's control.

[150] It is worth noting that Zimmerman holds that his view relates to moral responsibility (2002, pp. 554-555 ; 2015, p. 137), which he distinguishes from moral blameworthiness and deontic moral judgements. However, he holds that "responsibility is directly correlated, even if it is not identical, with susceptibility to reactive attitudes" (2002, p. 561), and so it seems unproblematic to discuss his account in terms of blameworthiness. In addition, Peels, who takes up and modifies Zimmerman's account, makes his case in reference to moral blameworthiness. (2015, p. 74) As I noted above, I take the common view that to be morally responsible for a negative outcome is to be blameworthy (see Nagel (1979) and Hartman (2017, 2019a, 2019b)).

Because the circumstances that one finds oneself in are often not in one's control, according to LT an agent cannot be more culpable than another simply because she faced circumstances that the other did not. Zimmerman holds that, despite the fact that it appears as if circumstances can affect one's degree of culpability, this is not the case if one appreciates the relevant counterfactual and evaluates the agent on the basis of it.

On the counterfactual view, if an agent S performs a morally bad action for which he is blameworthy, while another agent S' performs no such action but would have performed such an action if he had been in the same circumstances, S and S' are equally morally responsible. Zimmerman illustrates his view using the following example. George and Georg plan to kill an enemy. George successfully executes his plan to kill his foe, but Georg does not:

> Suppose that Georg, stung by Henrik's insults, had plotted long and hard for revenge, that he had repeatedly voiced his desire to see Henrik dead, that the day had come to put his plan into action, that he had positioned himself in the appropriate place at the appropriate time, that he had raised his gun and was on the brink of shooting point-blank at an unprotected Henrik — and that he was interrupted by a sneeze at the crucial moment.
>
> (Zimmerman 2002, p. 571)

But for the sneeze, Georg would have killed Henrik. Because the sneeze is a circumstance outside of Georg's control, if LT is correct then Georg's degree of moral responsibility must be the same as George's, even though he has not killed anyone. This is indeed the case, Zimmerman argues, because while the *scope* of Georg's blameworthiness is less than the scope of George's, the *degree* is the same. As the degree is what counts in determining one's blameworthiness, Georg and George are equally blameworthy. As Zimmerman puts it, "degree of responsibility counts for everything, scope for nothing… ." (2002, p. 568)

Zimmerman is aware that one may claim that it is unclear what exactly Georg is responsible *for*. He is also aware that this worry seems even more significant in other cases of apparent circumstantial moral luck. For instance, imagine an agent living a blameless life in a modern liberal democracy but who, if he had been born in Nazi Germany, would have committed all manner of atrocities.[151] Because the fact that he was born in a modern liberal democracy is a circumstance outside of his control, it seems as if on the counterfactual view he must be as blameworthy as an agent who was born in Nazi Germany and did commit atrocities. However, because he took no steps

[151] One might think that the type of person who would commit atrocities in Nazi Germany would also commit atrocities in any society in which he lived (albeit most of the time on a smaller scale.) However, as the example of Hanna from earlier in this chapter suggests, the way our personalities manifest themselves, and the morally relevant actions they influence us to take, is highly dependent on the society we live in.

to perform any morally bad actions — indeed, she never even conceived of the intention — it is not clear what she can be blameworthy for.

Zimmerman concedes that agents who would in other circumstances have done wrong (but do not in reality) are blameworthy for nothing. In his view, "there is nothing for which [such an agent] is culpable; his culpability has zero scope." (2015, p. 143) But this does not mean his degree of blameworthiness is zero; in fact, it is the same as that of the agent who was born in Nazi Germany and committed atrocities. An agent who is not blameworthy for anything (that is, whose moral responsibility has zero scope) can, on Zimmerman's view, nevertheless be morally responsible to a degree because he is morally responsible *in virtue of* some fact. In this case, he is morally responsible in virtue of the fact that he would have committed morally bad actions if he had been born in Nazi Germany. (Zimmerman 2000, pp. 143-144)

In summary, Zimmerman holds that the scope of one's moral responsibility can differ from the degree of one's moral responsibility. Although he does not put it as such, it appears that on his view scope is a measurement of what one is morally responsible for, whereas degree is a measurement of what one is responsible in virtue of. Because an agent can be responsible in virtue of being such that she would have committed a morally bad action, circumstances an agent faces cannot on Zimmerman's view affect one's degree of blameworthiness.

Having briefly laid out the counterfactual view, I am now in position to assess whether it is plausible. In what follows, I argue that it is not. I provide two arguments in support of this position. The first concerns the fact that the counterfactual view would seem to entail that agents are infinitely blameworthy; and the second relates to the fact that the view appears to sap blame of its communicative function.

One problem with the counterfactual view is that it seems to make agents blameworthy to an almost infinite extent. Zimmerman claims that an agent S who does nothing for which she is morally responsible (and therefore blameworthy) can nevertheless be blameworthy in virtue of the fact that she would have done *x* in situation *y* (where *x* is some morally wrong action and *y* is some particular set of circumstances). However, given that there is an infinite set of circumstances an agent might face in possible alternative worlds, under this infinite set of circumstances S would be infinitely morally responsible, and thus infinitely blameworthy.[152] This seems absurd. Note that this is true even of agents who would only ever commit very minor moral wrongs — telling a fairly inconsequential lie to their spouse, say, or parking in a disabled bay in a busy shopping centre carpark. Because the number of counterfactual scenarios is infinite, the degree of their moral

[152] Cf. Rivera-Lopez (2016) and Hartman (2017.)

responsibility is also infinite. This has the highly counterintuitive consequence that an agent who would only ever do minor moral wrongs is as blameworthy as notorious serial killer Ted Bundy (as both are infinitely morally responsible.)[153]

One way out of this would be to suggest that an agent could be constituted such that he would never do any wrong. It seems highly implausible to me that there is any agent who, no matter the circumstances, would not do wrong. But even if there are a few such agents, this does not save the counterfactual view from the charge of implausibility. This is because it would simply mean that the vast majority of agents were infinitely blameworthy, which is also intuitively highly implausible. Also, this view cleaves agents into two factions: one entirely morally pure and the other infinitely blameworthy, a dichotomous view of moral responsibility and blameworthiness which is perhaps even more implausible than the view that everyone is infinitely blameworthy. Therefore, I think we have good reasons to reject the counterfactual view.

Furthermore, as Hartman notes (2017, p. 67), such a view seems to imperil moral motivation. If one is morally responsible for what he would counterfactually have done, then only a tiny part of his degree of moral responsibility is determined by what he does in the actual world. If this is true, then it is unclear why we should care how agents behave in the actual world. This is highly counterintuitive.

Peels (2015) offers a way around this objection. He offers what he calls a modal solution to the problem of moral luck. On his view, moral philosophers have been insufficiently attentive to the nature of luck, and it is this lack of attentiveness that has prevented the development of a solution to the problem. As Peels, puts it, "the way out of this quandary is to remember that we are talking about luck." (2015, p. 79) Typically, Peels notes, philosophers have defined "luck" in the moral luck debate as existing "just in case it is beyond *S*'s control whether or not *E* occurs." (2015, p. 77) Peels holds that this definition is too broad and that a proper understanding of luck is as follows:

> An event *E* is lucky or unlucky for some person *S* at some time *t* iff (i) *S* lacks control over the occurrence of *E* at *t*, (ii) *E* is significant to *S* at *t*, and (iii) *E* occurs in the actual world, but does not occur in a wide class of nearby possible worlds.
>
> (2015, p. 77)

[153] Rivera-Lopez thinks that this argument from the absurdity of infinite blameworthiness shows only that constitutive moral luck must be possible: "To avoid this result [that agents are infinitely blameworthy], we have to say that there is a factor inherent to the agent, so that we can say of a particular agent that she would not have done an action *X* in any circumstance. But those essential features… are a matter of (constitutive) luck." (2016, p. 419) But, as I argue above, Rivera-Lopez is wrong. Even if an agent would never have done *x* in any circumstance, as long as there is some morally wrong action the agent would have done (even a very trivial one), then assuming there are an infinite number of possible circumstances on the counterfactual view the agent is infinitely blameworthy.

Lack of control over the occurrence of some event, then, is merely a *part* of luck. For instance, if one parks one's car somewhere and comes back later and finds that it has not been stolen, although this is outside of one's control and is significant to the agent, it does not count as lucky because one's car is not stolen in most nearby possible worlds. (2015, p. 78)

Because only things that are counterfactually true in nearby possible worlds count as lucky or unlucky, the problem of infinite blameworthiness evaporates. To illustrate this, Peels offers the example of three agents, Karl, Heinrich, and Alex. Karl and Heinrich both live in Germany during WWII. Heinrich finds out the hiding place of a Jew, and betrays him. His brother Karl would betray a Jew if he knew the hiding place of one, but because he does not have such knowledge he does not commit a such a betrayal. Alex lives in Australia sixty years after WWII and commits no comparable acts, but if he had lived in Germany during WWII and known the hiding place of a Jew, he would also have betrayed the Jew. According to Peels, the degree of blame of Heinrich and Karl is identical. However, Alex is not blameworthy, "because the world in which Alex does such a thing is far away and Alex's not betraying a Jew is, therefore, clearly *not* a matter of luck." (2015, p. 80) Agents are only blameworthy for being such that they would have performed morally blameworthy acts in nearby possible worlds, thus shrinking the degree of an agent's blameworthiness, and providing a response to Hartman's concern.[154]

But Peels provides no good reason to understand "luck" as used in the context of moral luck in this way. Consider Williams's comment that our ordinary understanding of morality is that it "require[s] a voluntariness that will be total and will cut through character and psychological or social determination, and allocate blame and responsibility on the ultimately fair basis of the agent's own contribution, no more and no less." (1985, p. 216) It is that the agent's control does not seem to be "total" that generates the problem; as Hartman notes, "the moral luck debate is not about luck per se but a tension in our ordinary thinking about moral responsibility." (2017, p. 24)[155] Therefore, even

[154] Note that Peels departs from Zimmerman here in regard to how an agent is counterfactually blameworthy. Zimmerman's view, as I noted above, is that an agent can be blameworthy *tout court* or *simpliciter*. That is, he can be blameworthy in virtue of the fact that he would have done some morally blameworthy act. Peels holds that this is "hard to make sense of," because "[t]he notion of blameworthiness simpliciter is not one we find in ordinary discourse, and as long as we have not been given a more detailed philosophical description of what blameworthiness *simpliciter* is supposed to be, [he finds] it hard to make sense of the idea." (2015, p. 76)

[155] See also Hartman (2017, p. 66), Hartman (2019b, pp. 3187-3188), and Statman, who notes the following: "it seems to make a rather minimal contribution to the question that is at the heart of what is known as the moral luck problem, namely, what role is played by what an agent actually does — in contrast to what she would have done based on her character, intentions, and efforts — in determining her moral status, in particular her praise- and blameworthiness." (2019, p. 199)

if Peels's analysis of the concept of luck is correct, it is unclear what relevance this has to the moral luck debate, given that the word "luck" as it has been used in this debate does not pick out the typical concept of luck. Rather, the "luck" in moral luck simply picks out a lack of control on the part of the agent.[156]

Peels's argument that one is only blameworthy for what one would do in close possible worlds is problematic for another reason. Hartman's argument from the communicative function of blame shows this. Hartman asks us to imagine an agent named Charles:

> Suppose that although Charles enjoys gambling, he has never been reckless. And in a close possible world but not in the actual world, Charles loses his job. This is devastating news for him, because the majority of his self-worth is bound up in that job. He attempts to distract himself from his newfound emptiness by heading to the closest casino. In that circumstance, Charles would freely make a series of reckless bets and lose his life savings. Suppose that this kind of action is out of character for Charles and that Jan, Charles's wife, knows that this counterfactual of freedom is true. Perhaps it is revealed to her by God or by an angel. In any case, she blames Charles by reacting toward him with resentment.
>
> (2017, pp. 67-68)

Hartman argues that blaming Charles for counterfactually gambling away the family's life savings lacks "communicative value" (2019, p. 230). This is because in blaming Charles, Jan presupposes that Charles is either theoretically or practically committed to the view that gambling away all his family's wealth is morally acceptable. However, Charles is neither theoretically nor practically committed to such a view. He is not theoretically committed to the view because he does not think that it is morally acceptable to gamble away his family's wealth. Nor is he practically committed to this view, because he takes no action in the actual world towards gambling away his family's wealth. Therefore, it seems absurd for Jan to blame Charles. But if it is absurd for Jan to blame Charles, then it appears that Charles is not blameworthy. And if Charles is not blameworthy for what he would have done in a non-actual world, then CML is possible.

In this subsection, I have briefly laid out the counterfactual view, and then argued that it has significant problems. It therefore does not provide us with good reasons to deny that there is circumstantial moral luck. The objection to my view, then, remains undefeated. In the following section, I consider another strategy to combat this objection: the view that there is extant circumstantial moral luck.

[156] For arguments to the contrary, see Hales (2015) and Lockhart and Lockhart (2017.)

7.2: Response 2: Circumstantial Moral Luck as Congruent with the Control Principle

I have argued in the preceding section that denying that there is extant circumstantial moral luck does not seem to be a fruitful strategy. In this section, I consider another way of rescuing my solution to the problem of unavoidable moral wrongdoing: conceding that circumstantial moral luck is possible, but holding that it is congruent with a plausible version of the Control Principle.

The view that resultant moral luck is impossible but circumstantial moral luck is possible has been dubbed by Hartman the "asymmetry thesis." If this view is defensible, and the Control Principle can be revised in light of this into a form that is supportive of the intuition RCW, then the solution I proposed to the problem of unavoidable moral wrongdoing dissolves. However, it is unclear that we have good reasons for considering the asymmetry view to be correct. After all, if luck in results is problematic in affecting one's degree of moral responsibility, then why is luck in circumstances acceptable?

I proceed as follows. First, I examine a compelling argument against the asymmetry view: Robert Hartman's parallelism objection. I argue that this objection does not go through, but that it is valuable in that its weaknesses highlight that not all of what is typically classified as circumstantial luck can affect an agent's degree of moral responsibility. His argument therefore points to a way of refining the Control Principle that allows for (some) circumstantial moral luck while still supporting the intuition RCW. Second, I make a positive argument for accepting circumstantial moral luck. I hold that Rivera-Lopez's argument that accepting circumstantial moral luck is necessary for the practice of holding agents responsible is, with a few tweaks, convincing; and this therefore provides us with additional reason for accepting a revised version of the Control Principle. As this version of the Control Principle supports the intuition RCW, this resolves this line of objection against my solution to the problem of unavoidable moral wrongdoing.

7.2.1: The Parallelism Argument

Robert Hartman's parallelism argument aims to show that circumstantial moral luck parallels situations of resultant moral luck in the relevant respects, and so accepting circumstantial moral luck necessitates accepting resultant moral luck, too. However, I argue that what Hartman's argument

actually shows is that while some types of luck in circumstances cannot affect an agent's degree of moral responsibility, this is not true of all types of circumstantial luck.[157] [158]

Hartman argues that important parallels between cases of circumstantial and resultant moral luck make it the case that resultant moral luck is possible. His argument is based upon a scenario involving three agents, and comparisons amongst those agents. He argues that because there is a difference in how blameworthy two agents are in one case (a case of circumstantial moral luck) there must also be a difference in how blameworthy two agents are in another case (a case of resultant moral luck). His argument is therefore designed to show that, if CML is possible, RML must also be possible. If this view is correct, then the asymmetry view — the view that CML exists but that RML does not — is false.

Here is the scenario Hartman uses to illustrate his point:

> Sneezy, the first assassin, is hired for murder but has bad allergies. When the time comes to pull the trigger, she suffers a sneezing fit. The fit renders her incapable of taking the shot. If, however, Sneezy were to have found herself in the same circumstance except that her allergies fail to be triggered, she would have freely taken the shot. Off-Target, the second assassin, has allergies just the same as Sneezy, but her allergies are not triggered. As a result, she has an opportunity and takes the shot. She, however, is off-target, because a bird catches the bullet. The comparative case of Sneezy and Off-Target is a standard example of circumstantial luck. They each would freely perform the same kind of morally significant action if they were in the same circumstance, but they do not have the same opportunities. Bullseye, the third assassin, has typical luck. Her aim is not obstructed by an allergic reaction, and nothing blocks the path of the bullet. She has an opportunity, fires a shot, and kills her mark. The case of Off-Target and Bullseye is a standard example of resultant luck, because they freely perform the same kind of action but with different results.
>
> (Hartman 2019a, p. 232)

Hartman holds that there are three parallels between the pairing of Sneezy and Off-Target and the pairing of Off-Target and Bullseye:

---

[157] Interestingly, in his book length treatment of the topic of moral luck, Hartman seems to endorse at least some of Wolf's argument in "The Moral of Moral Luck." In particular, he endorses her view that "different consequences [for which one is non-culpably responsible] can ground different moral requirements." (2017, p. 92) However, because he holds that there is extant resultant moral luck, Wolf's negligent driver would not be an example of an agent who faces such requirements, as the fact that the bad consequences result from negligence make the driver morally responsible for them.

[158] It is unclear whether Hartman's view on circumstantial moral luck allows for unavoidable moral wrongdoing. Hartman holds that the following principle circumscribes the effects of circumstances on moral responsibility: "An agent is not praiseworthy or blameworthy for an action or not fully so when features of her circumstance outside of her control either fully or partially damage a capacity that is necessary to exercise moral agency." (2017, p. 93) To illustrate this principle, he uses an example involving an agent who due to a family tragedy develops severe depression, resulting in an inability to fulfil a previous obligation. A "capacity" "necessary to exercise moral agency" seems to refer to something internal to the agent, and so it is plausible that Hartman's view does allow for extant unavoidable moral wrongdoing.

> [T]he agents in both case pairs (i) have identical agency in some sense and are saliently distinguished by luck, (ii) have very similar actual mental states, and (iii) bring about morally significant events that depend on their voluntary actions.
>
> (Hartman 2019a, p. 232)

From there, Hartman argues that because there is CML, and circumstances distinguish Sneezy from Off-Target, Off-Target is more blameworthy than Sneezy.[159] Furthermore, because the three agents are similar in the relevant respects, it follows that Bullseye is more blameworthy than Off-Target. It is unclear what justification one could have for thinking that, despite the fact that Off-Target is more blameworthy than Sneezy, Bullseye is not more blameworthy than Off-Target.

In order to analyse Hartman's argument, it will be useful to lay it out formally:

P1. There is circumstantial moral luck.

P2. Circumstances distinguish Sneezy from Off-Target.

C1. Therefore, Off-Target is more blameworthy than Sneezy (from P1, P2.)

P3. Sneezy stands in the same relation to Off-Target as Off-Target does to Bullseye.

Support:

a. The agents within each pairing have identical agency in some sense.

b. The agents within each pairing have similar actual mental states.

c. The agents within each pairing bring about morally significant events that depend on their voluntary actions.

---

[159] In addition to arguing that accounts that deny that there is extant circumstantial moral luck by using cases such as that of Ben (see below), Hartman argues in favour of circumstantial moral luck by analogy to evidential epistemic luck (first identified by Engel Jr. [1992]). Evidential epistemic luck occurs when it is a matter of luck that the agent has evidence justifying a proposition. Hartman holds that such luck is necessary for exercising epistemic agency, and that this is analogous to how circumstantial luck is necessary for exercising moral agency. (2017, pp. 95-103) This argument seems weak to me. Latus's point that "[w]hen it comes to deciding whether a person has knowledge or not, we do not concern ourselves with fairness" seems to highlight a relevant difference between the epistemic and moral domains. Hartman addresses the critique, but he misses the mark. He takes issue with Latus's statement that "If luck determines the moral status that a person has then how can it be fair for that person to have that status?" (2000, p. 166) Hartman argues that on his view luck does not "determine" one's moral status; it merely plays a role. But if one were to replace the word "determine" in Latus's sentence with "plays a role in," his point would still stand. It just does not seem as if luck is problematic in epistemic matters in the same way as in moral matters. Nevertheless, as I argue in 7.3.2, the fact that accepting some circumstantial moral luck is necessary for assigning moral responsibility does provide us with a reason to accept some forms of circumstantial moral luck. The problem for Hartman is that, as I argue below, the type of luck circumstantial luck Sneezy is subject to is not necessary for the practice of assigning moral responsibility.

P4. If the relationship between Sneezy and Off-Target parallels the relationship between Off-Target and Bullseye, then the blameworthiness of Bullseye relative to Off-Target also parallels the blameworthiness of Off-Target relative to Sneezy.

C2. Therefore, Bullseye is more blameworthy than Off-Target (from P3, P4.)

P5. Results distinguish Off-Target from Bullseye.

C3. Therefore, resultant moral luck is possible (from C2, P5.)

I turn now to analysing Hartman's argument. Let us first look at P3. I assume that Hartman identifies these features because he believes they are the relevant factors in determining moral responsibility. I will briefly explicate them. By "identical agency in some sense," Hartman means that either the agents actually have identical agency (in the case of Off-Target and Bullseye) or they have what he calls "subjunctively identical agency" (in the case of Sneezy and Off-Target), because Sneezy *would have* done exactly what Off-Target did, except that circumstances intervened. c) is somewhat vague, as it is not totally clear what morally significant event Sneezy brings about. I am inclined to agree with Nyman, who argues that what Hartman has in mind is the following: "in each comparison, the agent who brings about the morally worse event does so voluntarily (2022, p. 960).[160] b) — "similar mental states" — is straightforward.

Let us assume for the moment that this is what Hartman means. It is noteworthy here that, while a) and b) pick out similarities c) picks out a difference between the agents within each comparison, namely the difference in the badness of the morally significant events. Therefore, if these are the factors that are relevant for determining moral responsibility, and the agents are "identical" or "similar" in respect to a) and b), then if the agents within each comparison differ in their moral responsibility, this difference must supervene on c) — the fact that in each comparison, "the agent who brings about the morally worse event does so voluntarily." However, if the agent in each comparison is more blameworthy due to this factor, then Hartman begs the question. This is because, if Bullseye is more blameworthy than Off-Target due to the result of his action, then there is resultant moral luck. But this is the very thing that the parallelism argument is supposed to establish. If this is correct, then Hartman's parallelism argument fails.

[160] Hartman writes as follows: "[T]he event of taking the shot and the event of killing the mark both depend on the agency of the relevant person. In the case of Sneezy and Off-Target, the mere assassination attempt depends on Off-Target's voluntarily choice, and, in the case of Off-Target and Bullseye, the successful assassination depends on Bullseye's voluntary choice." (2017, p. 106-107) Thus, it seems as if what Hartman has in mind is that the morally worse event is the result of voluntary action.

It is of course possible that by c) Hartman simply means that each agent produces some morally significant events as a result of their voluntary actions (that is, he doesn't mean to imply that the difference in the badness of the outcome, produced voluntarily, is relevant.) But even if we assume this, Hartman's argument is not convincing. First, it is unclear what an agent's blameworthiness could supervene on. If the three agents are similar in all the three relevant respects, then blameworthiness must simply emerge from nowhere. Second, C1 is problematic. Hartman assumes there is extant circumstantial moral luck (P1), and so, because circumstance is what distinguishes Sneezy from Off-Target (P2), Off-Target is more blameworthy. Hartman claims this assumption is "dialectically benign," (2017, p. 107) because both he and advocates of the asymmetry view accept circumstantial moral luck. However, if we do not begin with the assumption that Off-Target is more blameworthy than Sneezy, there are at least some reasons to think that Off-Target *is not* more blameworthy than Sneezy. Consider the following facts of the case. According to Hartman's scenario, Sneezy is in position and ready to commit the murder when she sneezes, thus missing her chance. She has therefore (presumably) taken a whole range of actions designed to ensure the success of the assassination, such as procuring a weapon, assessing the proper position from which to take her shot, plotting her escape route, and so on. There is only one action Off-Target takes which she does not: pulling the trigger. It is unclear why this one action should render Sneezy less blameworthy. Indeed, when one holds all these facts in mind, it is plausible that Sneezy is just as blameworthy as Off-Target.

A slightly altered version of the Sneezy case brings out this intuition even more forcefully. Imagine a fourth assassin, Stuck. Stuck plans the killing, sees his target approaching, takes aim and so on in exactly the same fashion as Sneezy. However, he does not have a sneezing fit. Instead, when he goes to pull the trigger, he finds that it is stuck, and he is unable to get the shot away. His target therefore escapes unharmed. In this case, too, a factor outside the control of the agent prevents him from (fully) performing the action designed to result in the death of the target. However, it is highly counterintuitive that a stuck trigger can affect one's degree of blameworthiness, and so it is plausible that Stuck is just as blameworthy as Off-Target. Because it is not clear that there are any relevant differences between a stuck trigger and a sneeze, this gives further plausibility to the claim that Sneezy is just as blameworthy as Off-Target.

One can construct the story in such a way as to make it even more implausible that Stuck (and therefore Sneezy) is less blameworthy than Off-Target. Imagine that when Less Stuck (as we can call the assassin in this new construction of the story) pulls the trigger, it *does* move, only becoming stuck 0.01mm short of the point it needs to reach in order to effect the release of the bullet from the barrel

of the gun. In this case, Less Stuck has performed all the actions that Off-Target and Bullseye have performed except for moving the trigger the final 0.01mm. It is highly implausible that this can ground a difference in blameworthiness between Less Stuck and Off-Target.[161]

The reason C1 is problematic is not that either P1 or P2 is false. It is because C1 does not follow from P1 and P2. As Nyman notes in an illuminating critique of Hartman's argument, "[i]t does not follow … from the existence of circumstantial moral luck that it is present in all circumstantial luck scenarios." (2022, p. 960). That is, it is possible that it is only *in some cases* that circumstances outside the agent's control can affect blameworthiness. To help us understand when circumstances can affect moral responsibility, let us add yet another assassin to Hartman's collection, Would Have. Would Have spends months out of work but finally secures employment. However, had he not secured employment, he–driven by bitterness over his impecunious situation and a need to provide for his family–would have taken on the job of assassin and performed the murder. In this case, it seems absurd to hold that Would Have is as blameworthy as Sneezy, Off-Target, or Bullseye (for reasons I have outlined above; see 7.2.2). In this case, then, luck *does* play a role in one's degree of blameworthiness. Now consider again the scenarios involving Sneezy, Stuck and Less Stuck. These cases differ in an important respect from Would Have, in that these former three agents form an intention to do wrong, and take some steps towards doing it.[162] It is plausible that this distinction can ground treating luck differently in the case of Would Have from in the case of the other assassins.[163] Thus, although it is true that Sneezy, Off-Target, and Bullseye are similar in the relevant respects, it is not the case that Sneezy is less blameworthy than Off-Target, and so the parallelism argument does not show that if luck in circumstances can affect blameworthiness, then so can luck in results.

---

[161] It could be argued that the cases involving Stuck and Less Stuck differ from that of Sneezy, in that the former two agents will that the trigger be pulled, whereas Sneezy does not. On the view of Khoury (2018), for instance, what we are responsible for are "willings." As Stuck and Less Stuck will themselves to pull the trigger, but Sneezy does not, there is reason to hold that Stuck and Less Stuck are more morally responsible than Sneezy. But even if Khoury's view is correct it does not save Hartman's argument, because Khoury's view entails that there is no resultant moral luck. One is responsible for what one wills, not for what results from what one wills, and so results cannot affect moral responsibility. On this view, Sneezy is less blameworthy than Off-Target, but Off-Target is not less blameworthy than Bullseye.

[162] In fact, these cases *are* arguably cases of resultant moral luck. In the latter two, the agents apply an identical amount of force as Bullseye to an identical mechanism as Bullseye in the hope of effecting the same consequences as Bullseye; the only difference inheres in the outcome. Indeed, even the scenario involving Sneezy seems meaningfully different from paradigm cases of circumstantial moral luck, in that Sneezy develops an intention to commit murder and takes steps to effect the murder. Therefore, as Nelkin notes, it is unclear "what exactly counts as results… Is it anything after the beginning of the formation of the intention in cases when there is an intention?" (2019, p. 14) The line between circumstantial moral luck and resultant moral luck is therefore not as clear cut as has typically been assumed.

[163] See Nelkin (2019) and Nyman (2022) for an argument along these lines.

Indeed, as Nelkin notes (2019, p. 15 footnote), one of Hartman's arguments for the existence of circumstantial moral luck relies on a case like that of Would Have.[164] He asks us to imagine an agent Ben, who in the actual world is "an average person, neither very good nor bad." (2017, p. 66) However, in a nearby possible world, his parents die in a fiery aeroplane disaster as a result of errors on the part of the maintenance crew. In this counterfactual scenario, Ben murders several people connected to the airline out of a desire for vengeance. It is highly counterintuitive, Hartman holds, that actual Ben — who commits no murders — is as blameworthy as someone who does commit a series of vengeful killings.[165] But note that Ben and counterfactual Ben lack two of the three similarities — similar mental states and morally significant events depending on their voluntary actions — that Sneezy and Off-Target share and that Hartman holds are relevant in determining moral responsibility. In other words, he argues that circumstantial moral luck is plausible (in part) on the basis of one kind of scenario, and then uses another kind of scenario in his argument for the existence of resultant moral luck. But he provides no reason for us to think that these two scenarios are comparable.

If all this is correct, then the falsity of C1 means that Hartman's argument does not go through. Although an agent such as Sneezy is plausibly similar to Off-Target and Bullseye in the relevant respects, because there are no good reasons to think that he is less blameworthy than Off-Target, it is not the case that the parallel gives us a reason to think resultant moral luck is possible.

However, if this is right then not all circumstantial luck is circumstantial *moral* luck. I have suggested that Would Have is not blameworthy but Sneezy is as blameworthy as the other assassins. If so, it appears as if circumstances after the agent has developed a relevant mental state and taken some action related to this mental state cannot affect moral responsibility, but that circumstances that give rise to the mental state (and therefore the actions) in the first place can. Sneezy develops the intention to commit an assassination and takes actions to effect it, and so is just as blameworthy as Off-Target and Bullseye. However, Would Have does not develop such an intention, and (obviously) takes no actions to effect an assassination. Why is it, then, that some circumstantial luck can affect an agent's degree of moral responsibility, but not others?

It is plausible that what explains this difference is the following principle, which we can call the Opportunity Principle:

---

[164] Hartman uses a similar case elsewhere to illustrate the point that there are not relevant differences between resultant moral luck and circumstantial moral luck. (2019b, p. 3181)

[165] In addition to the intuitive plausibility of this, the arguments I made above against the counterfactual view support this view.

> **Opportunity Principle:** Once an agent has the opportunity to be morally responsible for *x*, factors outside her control cannot affect her degree of moral responsibility for *x*.[166]

On this view, luck in circumstances can establish opportunities for moral responsibility, but cannot affect an agent's degree of moral responsibility subsequently. Importantly, this principle supports the intuition RCW. This is because, if unavoidable moral wrongdoing were possible, it could be such that circumstances could affect an agent's degree of moral responsibility *without* this opportunity becoming available to the agent.

7.2.2: An Argument for (Some) Circumstantial Moral Luck

In the previous section, I argued that Hartman's initially compelling argument against the asymmetry view was flawed. In the course of my analysis, I noted that it was intuitively plausible that some types of circumstantial luck could affect one's degree of moral responsibility, while others could not. I ended by suggesting that the picture that emerged could be summed up in the Opportunity Principle. Here, I aim to buttress the plausibility of this picture by putting forth a positive argument for the asymmetry view. I do this by laying out an argument from Rivera-Lopez and showing how, with a couple of revisions, it can sustain this view.

Rivera-Lopez's argument in favour of the asymmetry view begins with the observation that it is circumstances that make moral responsibility possible in the first place. He argues as follows:

> Circumstances are necessary factors in making people responsible for their actions because they create the possibilities within which agents perform certain actions with specific purposes (or even negligently). On the contrary, events that directly affect the outcome of the agent's actions (or attempts) are not necessary to make the practice of responsibility possible. (2016, p. 422)

In order to illustrate his view, Rivera-Lopez asks us to imagine the following scenarios:

> Anne attempts to kill Robert and kills Robert in circumstance C1, for instance as Robert is walking on the street. Carla would attempt (and perhaps kill) Robert if he were in C1; but just two minutes before leaving, Robert decides to remain at home. Therefore, circumstance C1

[166] Nelkin defends a view similar to this: "[O]nce one has opportunities—and opportunities of a high-enough quality—then one is properly assessed only on the basis of what one does, or fails to do, with them. One's responsibility and degree of blameworthiness or praiseworthiness are a function of the quality of one's opportunity and what one does with it." (2019, p. 20) See also Nelkin (2013.)

> does not occur (C2 occurs, in which Robert is not available to be killed) and Carla does not attempt to kill Robert.
>
> (2016, p. 418)

According to Rivera-Lopez, C1 is necessary in order for Carla and Anne to be morally responsible, and therefore blameworthy. If Robert does not go out onto the street, then neither Carla nor Anne can attempt to kill him and so cannot be morally responsible for killing him. But this is not true of *the result* of an attempt to kill Robert. Both those who hold that resultant moral luck exists and those who do not would agree that, whether or not an attempt to kill Robert is successful, an agent who makes such an attempt has some degree of moral responsibility. Therefore, Rivera-Lopez argues, there is a relevant difference between circumstances and results (and, importantly, a difference that all parties in the moral luck debate can agree on): that circumstances are necessary for the existence of responsibility, whereas results are not. This difference, Rivera-Lopez holds, provides us with reason to believe that resultant moral luck is possible, whereas circumstantial moral luck is not.

One response to this line of argument is that C1 is not necessary for Carla to be morally responsible. Imagine that Carla plots her assassination of Robert carefully: she selects the optimal spot from which to shoot, purchases a gun suitable for the killing, engages in daily target practice to hone her accuracy, and so on. On the day she has picked out for the assassination, she even positions herself in her selected spot, ready to pick Robert off the moment he exits his apartment. Given all this, is it really the case that C1 — Robert's appearance on the street — is necessary for Carla to be morally responsible? It certainly seems plausible that it is not. It is tempting to hold that Carla is just as morally responsible as Anne, who makes the attempt on Robert's life; and it is intuitively very implausible that she is not morally responsible at all. If this is right, then it is not clear that Rivera-Lopez's claim about circumstances is correct.

Rivera-Lopez acknowledges this concern, but holds that, given that the moral luck debate centres around responsibility for what we do, and not for our character, there are reasons to discard such a position. Consider the following passage:

> Responsibility for what we do does not, for example, involve desires, pure intentions or plans that we may have made before (or independently of) performing an (intentional) action. Those mental events may carry some responsibility, but it would only be aretaic responsibility.
>
> (2016, p. 417)

Therefore, on Rivera-Lopez's view, non-aretaic moral responsibility supervenes on the performance of the morally bad act itself. It might seem as if this view begs the question against there being no circumstantial moral luck. If, in order to be morally responsible, one actually has to perform X, then,

given that the fact that the agent is in a position to perform X will at least in part be due to factors outside her control, either 1) circumstantial moral luck must be possible or 2) moral responsibility must be impossible. But Rivera-Lopez does not beg the question. Rather, this is simply another way of expressing his argument. That is, *because* we want to say that an agent can be morally responsible (in a non-aretaic sense), we ought to accept that circumstantial moral luck is possible.

I agree that non-aretaic responsibility has to supervene at least in part on action.[167] But this does not counter the argument I made above. After all, the planning that Rivera-Lopez dismisses as a locus of moral responsibility (except for in an aretaic sense) is itself very likely to involve action. For instance, (with reference to Rivera-Lopez's scenario) when Carla scopes out the neighbourhood, purchases a gun, and engages in target practice and so on she is performing actions. She is not performing *all* the actions that Anne performs — she does not pull the trigger of the gun when she sees Robert on the street, because Robert does not go out onto the street. But she is performing some actions, and Rivera-Lopez gives us no reason to think that these actions cannot ground non-aretaic responsibility, and make Carla blameworthy to the same degree as Anne.

But it is possible to reconstruct Rivera-Lopez's argument so that it evades this problem. Let's say that in the Carla case it is the fact that Robert spurns Carla's romantic advances that leads her to attempt to kill him. Now imagine that Carla had never met Robert. If Carla had never met Robert, then she would never make any romantic advances towards him for him to spurn, and so she would never actively plan to kill him. If she never actively planned to kill him, then it is unclear how she could be morally responsible to any degree. Therefore, the circumstances of Carla meeting Robert and Robert rejecting her advances are necessary for the practice of assigning moral responsibility. And so Rivera-Lopez is correct in holding that there is a relevant difference between circumstances and outcomes, and we therefore have a reason for accepting some circumstantial moral luck but denying resultant moral luck.

Robert Hartman counters that, even if it is true that accepting circumstantial moral luck is necessary for assigning moral responsibility and resultant moral luck is not, this does not provide us

---

[167] On Khoury's view (2018) that what we are morally responsible for (in a non-aretaic sense) is "willings," action and what we are non-aretaically morally responsible for can come apart. Khoury shows this by asking us to imagine that an assassin who believes that he positions himself for the murder, takes aim at the target, and pulls the trigger. However, unbeknownst to him, his brain has been extracted, placed in a vat, and hooked up to a computer, and his experiences take place within a virtual environment. Khoury holds that this shows that we can be morally responsible without acting. On this view, an agent such as Sneezy would indeed be less blameworthy than Off-Target (because he did not will the pulling of the trigger.) However, even if this view is correct it still supports the asymmetry view. This is because circumstances outside of one's control can play a role in determining what one wills and so can affect an agent's degree of moral responsibility, but results cannot play such a role, and so are irrelevant to one's degree of moral responsibility.

with a good reason for accepting circumstantial moral luck and rejecting resultant moral luck. Hartman argues that Rivera-Lopez merely asserts the principle that we should accept moral luck only if it is necessary to make possible the assignment of moral responsibility (Hartman dubs this the "Relevant Conditional"); he provides us with no argument for why we should accept it. Nor can we rely on intuition, Hartman claims; "plausibly, the intuition that motivates the Relevant Conditional is the same intuition that motivates the Asymmetry View itself," (Hartman 2017, p. 109) and so this would beg the question in favour of the asymmetry view. Furthermore, he questions why we should "accept only the bare minimum of moral luck required to be morally responsible agents… . [I]f we accept circumstantial moral luck to be morally responsible for actions, why not also accept resultant moral luck to be additionally morally responsible for consequences?" (2017, p. 108) In other words, Hartman agrees that circumstantial moral luck is necessary for moral responsibility to be possible, but sees no reason why we should accept moral luck only if it is necessary in this way.

But the force of Hartman's claim that the Relevant Conditional simply reasserts a fundamental anti-moral luck intuition is deadened when we consider that the problem of moral luck generally stems from two conflicting intuitions: 1) the abstract intuition that the degree to which we are morally responsible cannot depend on factors outside our control (the Control Principle), and 2) the intuition that arises when we consider specific situations that factors outside our control *do* affect the degree to which we are morally responsible. The Relevant Conditional is therefore the product of a process of reflective intuitionism seeking to reconcile the two intuitions that together comprise the problem. Rivera-Lopez's Relevant Conditional brings these two intuitions into equilibrium by modifying both 1 and 2 such that they cohere. This explains why we do not also accept resultant moral luck — because doing so would represent an abandonment of the Control Principle, whereas Rivera-Lopez's Relevant Conditional only requires that we modify it. Therefore, absent good arguments in favour of abandoning the Control Principle rather than merely modifying it to bring it into coherence with 2, Rivera-Lopez's solution seems preferable.

What Hartman calls the Relevant Conditional, Rivera-Lopez himself terms the Main Reason. He expresses it as follows:

> **Main Reason:** The occurrence of fully uncontrollable events that are not necessary for the existence of responsibility cannot affect the degree of responsibility.

Note that this is substantively identical to the Opportunity Principle (OP) I articulated above:

> **Opportunity Principle:** Once an agent has the opportunity to be morally responsible, factors outside her control cannot affect her degree of moral responsibility.

These "fully uncontrollable events … not necessary for the existence of responsibility" are the opportunities for being morally responsible I refer to in OP. Therefore, given that as I have shown Rivera-Lopez's view largely succeeds, we have strong grounds for considering OP to be correct.

In this section, I have argued that, given the method of reflective intuitionism, there are reasons to accept revisions to the Control Principle in line with the Relevant Conditional. If we wish to assign moral responsibility to agents, then (some) circumstantial luck is necessary, whereas results are not. Contrary to Hartman's claims, accepting this limited role for luck is not reasserting the Control Principle. Rather, it is the product of a revision in order to bring two conflicting intuitions into a state of equilibrium. We therefore have good reasons to accept some circumstantial moral luck, and so the problem that I identified at the start of this chapter is dissolved.

7.3: Three Further Objections

In this section, I consider some objections to the asymmetry view I have argued for above. I begin with an objection that there is a contradiction in my view, in that I am holding that situations of UMW and situations of RML are both analogous and disanalogous. Then, I consider the objection that allowing for circumstantial moral luck would seem to allow that one ethnic group could be more blameworthy than another. I reject all three objections.

7.3.1: Objection from the Recognition of Moral Error

One objection which could be levelled against the view of moral responsibility I defend above is that it cannot account for cases where an agent forms an intention and takes some meaningful action towards making the intention a reality, but then decides against proceeding as a result of recognising her moral error.

Here is a scenario to illustrate this objection. Imagine that an agent Susan knows that her husband's life is insured for a large sum and so forms an intention to kill him to receive the lucrative payout. Susan makes some preparations — purchases a gun, arranges a false alibi, etc. — but then is

struck by the immorality of what she is doing, and cancels her murderous plans. In this case, we are tempted to say that Susan is not as blameworthy as one who actually goes through with the murder.

But now imagine a scenario in which another assassin, Beth, plans to kill her husband. Everything is the same as in the previous scenario, except that, the day before she would have reneged on her plans to kill her husband, Beth is hit by a truck as she is crossing the street and dies. She therefore does not have the opportunity to renege on her plans and so is as blameworthy as an agent who goes through with the murder. But note that the fact that the truck runs her down and kills her is out of her control; and so it appears that the fact that she is as blameworthy as one who actually commits the murder is a case of moral luck. However, according to the revised version of the Control Principle I stated above, factors outside of an agent's control can only play a role in establishing an agent's moral opportunities; once these opportunities are established luck can play no role. But if the fact that Beth is run down by a truck — a fact that is outside of her control — affects her degree of blameworthiness, then these factors *can* affect blameworthiness. We have therefore arrived at a contradiction.

One way to attempt to overcome this contradiction is as follows. One could argue that the fact that Beth would, counterfactually, have reneged on her evil scheme had she not been run over by a truck makes her no more blameworthy than Susan. This is however an unsatisfactory solution. To see why, imagine that Susan only decides not to proceed with the murder after her sister gives her a copy of the play *Macbeth* and she reads it and experiences a moral epiphany regarding the wrongness of murder. However, if Susan's sister had not given her a copy of *Macbeth* she would have proceeded with her plans and murdered her husband for the life insurance pay out. Therefore, this solution seems to allow for moral luck after the initial opportunity and therefore contradicts the Opportunity Principle I stated above. In other words, luck does not affect their degree of blameworthiness.

A more promising means of resolving this apparent contradiction evades this problem. On this view, Susan is just as blameworthy as Beth for having the intention to murder and taking some action to realise the intention. However, Susan is also praiseworthy for recognising the error of her ways and reneging on her plans to kill her husband.[168] This of course assumes that praise- and blameworthiness do not cancel each other out; but this is a position held by those on both sides of the moral luck debate, and I think a fairly plausible one.[169]

---

[168] Does this mean that Susan is more deserving of punishment than Beth? On a purely retributive model, the answer seems to be yes. However, it is plausible to me that punishment can come apart from blame due to consequential concerns. If this is correct, then it could be that a greater punishment for S is warranted due to the fact that, because S has not recognised that she is in moral error, she is more likely to commit a similar act of wrongdoing in the future.

[169] Cf. Hartman (2017) and Zimmerman (2019).

If this is correct, then we have a way of accounting for cases in which the agent forms an intention to commit moral wrong, but then reneges on the intention.

7.3.2: The Contradiction Objection

Another possible objection to my view concerns my argument that situations of RML and UMW are analogous. In the previous chapter (see 6.6), I argued that although the position of the luck was different in situations of UMW from in situations of RML, this was irrelevant. In neither type of situation, I argued, was the agent (fully) culpable for the harm she caused because the harm was the product of factors outside the agent's control.[170] It is for this reason that both situations could generate agent-regret. But if these situations are analogous — if, as I argued, the position of the luck is irrelevant — then how can the asymmetry view be correct? It seems that I am arguing that the position of the luck is both relevant and irrelevant, which is incoherent. Can the arguments I made in the previous chapter be reconciled with the asymmetry view?

It is true that the asymmetry view requires that an agent in a situation of UMW has a (possibly) difficult moral choice; that this choice will likely result in moral distress; and that it will give rise to subsequent moral requirements. An agent who faces a difficult moral situation and makes the morally wrong choice (let's say for self-serving reasons) cannot claim that she is no more blameworthy than someone who did not face a difficult situation of moral conflict but, if she had, would also have made the wrong choice for self-serving reasons. Similarly, even if such an agent acts in the morally correct way but does not fulfil the subsequent moral requirements that arise, she cannot claim that she is no more blameworthy than an agent who would also have failed to fulfil these requirements but who was not causally responsible for any harm.

In this way, situations of UMW *can* give rise to different moral evaluations, and in this respect they are different from situations of RML — I have argued that there is not extant RML, and so factors affecting the outcome that are outside the agent's control cannot affect the agent's moral evaluation. So in this respect, situations of UMW and RML are disanalogous.

But this does not conflict with my claim that the situations are analogous, because they are analogous in another respect: in respect of the features that give rise to the requirement to take responsibility for non-culpable harm. In cases of UMW, an agent is torn between two obligations.

[170] In cases of UMW, the agent is not culpable for the harm. In cases of resultant moral luck, I argued that the agent was to some degree responsible for the harm, but no more responsible than an agent who performs exactly the same action but does not cause harm. Exactly how responsible an agent such as Driver_1 is is a difficult issue, as Wolf (2001) notes.

Even when she takes the all-things-considered morally right choice action, the fact that there is an obligation not acted upon results in harm. The agent is therefore causally (but not morally) responsible for harm. When responding to moral reasons, then, a rational agent lacks control over the production of bad consequences. This is analogous to an agent in a situation of RML. Such an agent makes a choice that is not all-things-considered morally correct, but whose consequences are worse than seem warranted by the blameworthiness of the action. While the agent is not properly responsive to moral reasons — his choice is all-things-considered incorrect — he is responsive to moral reasons to the same extent as one who acts identically in identical circumstances but whose actions produce no bad consequences. Thus, in both cases there is harm that extends beyond the agent's moral culpability. In this way, situations of UMW and RML are analogous.

The way in which the situations are analogous, then, supervenes on different features of the situations from the way in which they are disanalogous. The charge of incoherence is therefore baseless.

7.3.3: Objection from Group Differences in Blameworthiness

Another possible objection to the idea that there is extant circumstantial moral luck relates to the ostensible implausibility of the idea that one group of people (particularly an ethnic, racial, or religious group) can be more blameworthy than another group due to factors outside the control of the members of the group. However, accepting circumstantial moral luck as I have done seems to mean that this is indeed the case.

Above, I have argued that there are good reasons to accept that circumstances outside the control of the agent can affect how blameworthy the agent is. But this leads to the following problem: if it is unproblematic for circumstances to affect the blameworthiness of a single agent, then it must be unproblematic for circumstances to affect the blameworthiness of many agents. And if it is unproblematic for circumstances to affect the blameworthiness of many agents, it is unproblematic for circumstances to affect the blameworthiness of many agents who are all united by some common feature. But it seems that there are situations in which we do not think this. Consider the following situation (it is hypothetical, although of course it closely reflects the reality in many countries.) In country Y, there are two ethnic groups, A and B. For whatever reason group A on average faces more economic and social hardship than group B, and therefore faces more challenging morally relevant circumstances. Perhaps the paucity of economic opportunities means the members of group A more often face circumstances conducive to selling drugs which lead to widespread addiction and harm,

for instance, or to form or join criminal gangs which inflict violence upon the community. Because of this, A does more wrong than B, and is more blameworthy.

That members of one ethnic group can be more blameworthy, on average, than members of another due to the circumstances they face is a contention that most would wish to resist. But if the arguments I made above about CML are correct, then it is unclear how we can resist it.

One possible response to this objection is that what we find objectionable is not the fact that members of group A are more blameworthy than members of group B, but rather the injustice that produced the circumstances in the first place. For instance, perhaps previous generations of group B were responsible for crafting discriminatory laws and social practices and customs, and so the present members of group B have a responsibility to ameliorate inequalities that stem from these laws, practices and customs. In the United States, for example, the practice of slavery robbed people of African descent of a whole range of possibilities and opportunities for development and flourishing. Because the institution of slavery was unjust, and this institution is responsible (at least in part) for the more difficult circumstances African Americans face in present day US society, we find these circumstances problematic. If this view is correct, then the fact that the members of group A are more blameworthy is true, but the circumstances which give rise to this blameworthiness are the product of injustice, and this is what we find problematic.

One could however retort that this response is unsatisfactory. Imagine again country Y. If circumstantial moral luck is possible, then the following statement is true:

> **Statement 1:** *Members of group A are, on average, more blameworthy than the members of group B, and this is due entirely to circumstantial luck.*

The statement says nothing about unjust social, legal, or cultural practices producing the circumstances. However, when I consider this statement, I have the intuition that it is problematic. It simply does not seem that it can be correct that members of one group can be more blameworthy than members of another due to luck. I suspect that many would agree with me. Therefore, it seems as if it is not simply the unjust social circumstances which produce the greater blameworthiness — and therefore wrongdoing — that are problematic. It appears that the greater blameworthiness itself is also problematic. But if this is correct, then circumstantial moral luck cannot exist. This is because there is no reason to think that circumstances outside a single agent's control can affect his degree of blameworthiness, but that circumstances outside the control of a group of agents united by some common identity cannot affect their degree of blameworthiness.

But there are reasons to think that such an intuition is the product of bias. In order to show how, let us pretend that we somehow acquire the knowledge that people born in the year 1976 are, on average, more blameworthy than people born in any other year, and that this is due entirely to CML (perhaps an angel descends from heaven to inform us of this.) If I were to discover this, I would certainly find it interesting and perhaps be curious as to what the circumstances were that gave rise to this greater blameworthiness. However, I would not find it problematic. But this suggests that what seems intuitively wrong about Statement 1 above is not that there is an average difference in blameworthiness between members of the two groups, but rather that the difference is a difference between members of ethnic groups.

Why should we be particularly sensitive to statements about average differences in blameworthiness among ethnic groups? Why should they seem problematic in ways that statements about average differences among age cohorts do not? It seems to me that this has to do with the likely consequences of such statements. Because of the history of racial discrimination in the Western world, and the terrible consequences this discrimination has had, we are likely to be especially sensitive to statements that paint one ethnic group (particularly an historically disadvantaged one) in a negative light, and to worry that issuing such statements will result in renewed discrimination (through igniting old prejudices or providing fuel for fringe racial supremacist groups or whatever.) We therefore have a *prima facie* obligation to refrain from issuing such statements (even if they are true.)

This argument is strengthened when we consider another statement:

> **Statement 2:** *Members of group B are, on average, less blameworthy than the members of group A, and this is due entirely to circumstantial luck.*

When I consider this statement, I do not have the same intuitive aversion to it as I do to statement 1. This presumably because, as I stipulated above, group B is economically and socially advantaged relative to group A, and so I am not worried that the issuing of the statement will result in discrimination or persecution or any other negative consequence.

In other words, because there is extant circumstantial moral luck, we are epistemically justified in believing the proposition expressed in Statement 1. However, we are unjustified in *acting* upon this belief (even if this action is confined to the mere expression of the proposition.) This may seem contradictory; it may seem that, if it is true that *p*, then *p* provides reasons for action and so assuming we ought to act in accordance with reasons then we ought to act in accordance with *p*. But

this ignores the fact that *p* is simply one reason for action amongst many; and if it is outweighed by other reasons, then it is unproblematic not to act in accordance with it. In the case of Statement 1, those other reasons — such as that making statements about group A's blameworthiness will likely produce morally bad consequences — outweigh *p*. It is therefore not irrational to avoid acting upon *p.* [171]

My aim here has been to sketch out a possible response to the objection that allowing for CML would lead to the problematic conclusion that certain ethnic groups can be more blameworthy than others. This is obviously a topic on which much more could be said; offering a more thorough consideration of it would take me too far from the main focus of this thesis. However, the argument above at least offers a plausible means of reconciling our aversion to statements about the blameworthiness of particular ethnic groups with my contention that circumstantial moral luck exists.

Conclusion

In this chapter, I considered two strategies for showing that my solution to the problem of unavoidable moral wrongdoing was not undermined by the fact that it apparently required that there be extant circumstantial moral luck. I began by examining the possibility of denying that there is extant circumstantial moral luck (contrary to appearances.) I first considered Enoch's argument for why circumstantial moral luck does not exist. Enoch argued that, while there is what he calls "morally

---

[171] My solution here is akin to what Basu, in the context of the moral encroachment debate, calls "downstream accounts of the wrong." (2018, p. 2501) That is, the wrong is not in the belief, but is in the (downstream) action that results from the belief. Basu is critical of such accounts, offering the story of a racist hermit who "will never interact with the disadvantaged person he believes something negative of, [and] he will never interact with or contribute to the institutional structures of racism." The hermit "comes to believe that Sanjeev smells like curry ... [because] he discovered some trash on the ground which happened to be an alumni newsletter from Sanjeev's university that included a picture of him." Basu adds that "Sanjeev happens to have recently made curry so in this instance the hermit's belief is true — Sanjeev *does* smell of curry." Basu argues that even though the hermit will not commit any wrong action on account of his racist view, it is still wrong for him to hold this view. This, Basu argues, shows that the "downstream view" is incorrect. But Basu apparently fails to appreciate that there is an important difference between the hermit's view and the view that (to use another example from Basu) a black man in the lobby of an exclusive club must be an attendant (there are very few black members and almost all the attendants are black.) In the latter case, there is substantial relevant evidence in the form of statistics that the man is an attendant. In the case of the racist hermit, there is no evidence for his belief, which is only correct through chance. Basu's example, then, does little to dispute downstream accounts.

significant plain luck," — luck that determines our obligations — this is distinct from circumstantial moral luck, or luck which shapes our moral appraisals. This is because an agent is appraised not simply on what he does, but on how he plays his moral hand. I countered that Enoch's argument had significant weaknesses, and at best ultimately collapsed into another anti-moral luck view advanced by Zimmerman: the counterfactual view.

It is this view that I addressed in the next section. Zimmerman's counterfactual view holds that if an agent does not do morally wrong action *x* but would have done *x* had she been in circumstance *y*, then she is just as blameworthy as an agent who is in circumstance *y* and does *x*. I argue that Zimmerman's view, if one accepts it, leads one to some very counterintuitive conclusions regarding blameworthiness. In particular, it appears that everyone is infinitely blameworthy under such a view, and that blame also loses its communicative function. I argued that these implausibilities constitute good grounds for rejecting his view.

Then I moved on to analysing the second strategy: the asymmetry view. I argued that there are good reasons to accept that circumstantial moral luck differs from resultant moral luck in an important respect: that circumstantial moral luck is necessary for the practice for assigning moral responsibility, whereas results are not. I contended that Rivera-Lopez's argument (2016) along these lines was successful, subject to a few modifications; and there is therefore a reason to accept some circumstantial moral luck but not resultant moral luck.

Chapter Eight: Responsibility and Obligation

Introduction

In the previous chapter, I argued that the fact that Susan Wolf's solution to the problem of resultant moral luck seemed to require that there be extant circumstantial moral luck was unproblematic. In doing so, I addressed what appeared to be a serious threat to my view: that it seemed to require that circumstantial moral luck be possible. In this chapter, I want to address another aspect of my view which I have so far explored only briefly (see 5.1): the idea that an agent who has faced a situation of unavoidable moral wrongdoing can be required to take responsibility for harm that arises as a result of her all-things-considered morally right action. At first glance, this idea appears puzzling; and Wolf provides very little detail on what it involves. In what sense does one take responsibility? What does such responsibility involve? How can we determine what obligations arise from taking on such responsibility?[172]

My aim in this chapter is to sketch out an answer to such questions. I begin by revising to some extent Wolf's account of what the nature of our requirements is in situations in which we have caused harm despite performing an all-things-considered morally right action. Then, adapting Sommers's "negotiating responsibility" view of moral responsibility, I offer an account of how an agent's responsibility for such types of harm translates into a set of obligations. To end, I address a couple of possible objections to my argument.

Finally, it is worth noting that while much of what I say is, I think, applicable to situations in which an agent is responsible for some sort of harm which he is not morally blameworthy for, I focus mainly on situations of apparent unavoidable moral wrongdoing, and the harm that results from an agent's all-things-considered morally right actions in such situations.

8.1: Requirement to Take Responsibility or Direct Vicarious Responsibility?

In this section, I aim to refine slightly the Requirement to Take Responsibility (RTR) view I defended in Chapters Four to Six: that we can be required to take responsibility for what we are merely causally responsible for. In particular, I argue that, in such situations, it is implausible that there is a duty for

[172] In this chapter, I focus on what an agent's obligations are to a party who she has non-culpably inflicted harm upon. Payson (2009), drawing on Ruth Marcus's contribution to the moral dilemmas debate (1980), makes the interesting argument that such an agent might also have an obligation to work to alter societal structures so that moral conflicts do not arise in the future. Whether or not this is correct, I set it aside in this chapter.

us to take responsibility. Rather, it is that we are directly responsible simply in virtue of the harm we have caused. No act of responsibility-taking is necessary in order to face the moral requirements stemming from the harm.

To argue for this position, I will begin by sketching out an argument from David Enoch (2012). Enoch defends and elaborates on Wolf's position: that there can be a duty to take responsibility, and it is the performing of this duty — that is, the taking of responsibility — that generates the subsequent moral requirements. After I have laid out his view, I will argue that it is problematic in several respects, and that a better explanation is that an agent faces certain moral requirements simply in virtue of the fact that she is causally responsible for harm. In taking this line, I endorse the view of Goetze and others,[173] who have held that there are situations in which we are directly responsible for harms for which we are not morally responsible.

8.1.1: Duty to Take Responsibility: Enoch's Account

In this section, I lay out the view of David Enoch, which expands on the view that Wolf gestures towards in her article.

On Enoch's view, it is plausible that there are situations in which one can be required to take responsibility for some bad consequence *x*,[174] even when one is not morally responsible for *x*.[175] He contends that this view best accounts for the intuitions that arise in certain cases. He begins his argument by giving an analogy to the practice of making promises. When one promises to bring "the salad [to a picnic], what [one is] doing seems to be to undertake a commitment to bring the salad."

---

[173] See also Mellor (2021.)

[174] For Enoch, these include situations in which the agent, as a result of making an all things considered morally correct choice, nevertheless causes harm; but they are not limited to such situations. Other situations in which an agent might be obligated to take responsibility include when she causes harm non-culpably (such as Williams's truck driver does); and when one's child or one's country has committed some sort of wrongdoing. As I stated in the introduction, I am concerned with what an agent's obligations are in the wake of his facing a situation of apparent unavoidable moral wrongdoing. However, because the scenarios Enoch uses to illustrate his point often involve other types of non-culpable harm, I discuss these types of scenarios in showing that his view is implausible.

[175] It should be noted that Enoch is somewhat restrained in the nature of his claims: "I take myself here to establish more a possibility-result than an outright conclusion. Thus I shall not defend the claims that in these cases you are not responsible and that you are rather under a moral duty to take responsibility. Rather, I shall defend the claim that this is a coherent conjunction, and furthermore a potentially productive one, one that seems to capture something dear to our heart in the phenomenology of responsibility." (2012, p. 100) In other words, Enoch seeks to establish that this conjunction is merely logically possible. It seems to me that Enoch attempts to do somewhat more than this; he appears to argue that his view is plausible rather than merely possible. In any case, my analysis of Enoch's account here aims to show that there is a better explanation.

The promise therefore generates the subsequent obligation to bring the salad, as "the making of a promise is itself an action," and so one "may be morally required to make a promise." (2012, p. 105) Enoch notes that there is an obvious difference between this and situations in which one is required to take responsibility: it does not seem as if one is *required* to promise to bring the salad. However, Enoch holds that there are situations in which one is required to make promises:

> [A] single parent friend of mine is diagnosed with severe illness. He is extremely anxious about who will take care of his daughter if he dies. Suppose that we are not close enough for it to be morally required that I take care of her. Indeed, suppose that if my friend dies without anyone having committed to taking care of his daughter, still one or two closer friends will step up to the plate. But suppose further that none of them will *promise* my friend now that they will. By promising, I can significantly alleviate my friend's anxiety. It seems to me that there are ways of filling in the details of the example such that I shall be morally required to promise my friend to take care of his daughter if he dies.
>
> (2012, p. 105)

In such a situation, Enoch argues, one is not required (prior to any act of responsibility-taking) to take care of his friend's daughter. However, one *is* required to take on responsibility for caring for her. Furthermore, once one has made this promise , he is responsible for taking care of her. This shows that it is conceptually unproblematic for there to be a requirement to take on some responsibility prior to there being a requirement to perform the actions that are involved in taking on this responsibility. Although Enoch acknowledges that is not an example of a situation in which prior harm generates the duty, it shows that it is logically coherent for there to be a duty to take responsibility.

Enoch argues that the plausibility of this is strengthened when we consider the case of Williams's non-negligent driver who injures a pedestrian. He contends that if his theory is correct, then there is a difference between a driver who non-culpably injures a pedestrian and visits her in the hospital once and never again, and one who never visits the injured pedestrian at all. The driver who visits once takes responsibility by visiting, and so her moral failing is to fail to fulfil the requirements entailed by taking responsibility. In contrast, the moral failing of the driver who never visits at all is that she fails to take responsibility in the first place. Enoch holds that "[i]ntuitively, this sounds exactly right… . [T]here is a difference between the two flaws involved in these two kinds of behavior." (2012, p. 111) Because Enoch's account captures our intuitions about this difference, we have strong grounds to consider it correct.

Enoch argues that the responsibility that an agent can have a duty to take is not (only) a forward-looking responsibility. Rather, it is largely backward-looking — one is taking responsibility

for something that has happened, not just accepting certain responsibilities going forward.[176] However, this responsibility is not to be conflated with moral responsibility, as the agent has (*ex hypothesi*) not done wrong. Enoch, building on the view of Oshana (1997) characterises the type of responsibility that such an agent can be required to take as follows:

> *A is responsible (in this sense) for X to P if and only if, if X is … morally problematic[177], A owes it to P to justify X (or the relevant X-related thing), or to offer an excuse for it, or to apologize for it, or to explain it, or something of this sort.*
>
> (2012, p. 118, italics in original)

In other words, when an agent A commits some act of non-culpable harm (this is what Enoch seems to mean by an "otherwise morally problematic" action) he is responsible for it when he owes it to harmed party A to somehow account for his actions. I will refer to this type of responsibility as non-culpable responsibility.

Enoch is aware that, if this is the sense of responsibility that applies to agents who have non-culpably committed harm, then one could deny that such agents had a duty to take responsibility, but were not responsible until they had done so. This is because the act of apologising is thought by most to be a "paradigmatic way of taking responsibility" (Enoch 2012, p. 119). But if one has to be responsible in order to owe an apology, then the agent must be responsible simply as a result of causing the non-culpable harm. If this were not the case, then one would not owe an apology in the first place. Enoch has several responses to this. First, he notes that "while apologizing may in a sense be the paradigmatic way of taking responsibility, it is not the only one." (2012, p. 119) Second, he argues that the directionality of the duty to take responsibility may come apart from the directionality of the apology. For instance, it could be the case that

> a father owes *his son* the duty to take responsibility for (some of) the son's actions, but owes *the son's crime's victim* an apology as a constitutive part of his being so responsible. If so, the duty to take responsibility (which may come to a duty owed *to the son* to apologize to the victim) and the responsibility taken (which may come to a duty owed *to the victim* to apologize to him) are not after all identical.
>
> (2012, p. 119. Italics in original)

---

[176] Enoch notes that this distinguishes it from responsibilities that arise as a result of promise making (see pp. 104-105 .)

[177] Enoch here writes that "*X is wrong or otherwise morally problematic*… ." However, given that he is talking about responsibility that arises in cases where there is no moral failure, it is unclear why it is necessary to include "wrong" in the definition. Perhaps he means "wrong in normal circumstances." Even if this is true, however, it adds nothing to the definition.

Third, Enoch argues that even if an apology is the correct way to take responsibility, and even if the apology and the duty to take responsibility are owed to the same person, this does not undermine his point. This is because "the duty to take responsibility is more basic here, and … the duty to apologize is derivative—one only has the duty to apologize in virtue of having a duty to take responsibility." (2012, p. 119) By this Enoch seems to mean that the fact that one has the intuition that one ought to apologise is really picking out the fact that one has a duty to take responsibility, which will then generate a duty to apologise.

In summary, Enoch holds that it is perfectly coherent to contend that an agent can have a duty to take responsibility for some bad outcome, action, or event for which she is not morally responsible. In taking this line, he seems in accordance with Wolf's view. In the next section, I'll argue that we have good reasons to reject this view, and instead to adopt a view that, in the situations Enoch describes, we simply *are* responsible for harm; there is no need to take responsibility.

8.1.2: Problems with Enoch's Account

As Enoch acknowledges, his view is a justification of and elaboration on Wolf's view. As I appeared to commit myself to this view earlier (see Chapters Four-Six), it might be expected that I would endorse his view. But I do not. In this section, I suggest that a better explanation is that our obligations arise in virtue of being directly responsible for the bad state of affairs that arises.

The most obvious problem is as follows. Plausibly, if there is a requirement for S to do *x*, and doing *x* entails doing *y* and *z,* then there is a requirement for S to do *y* and *z.* For instance, if I am required to visit my mother to celebrate her birthday, and I know that such a visit will involve going out for a family dinner, then I am required to go out for a family dinner. But if this is the case, then an agent who is required to take responsibility for bad consequences she has non-culpably caused is also required to perform the actions that taking responsibility for these bad consequences entails. Therefore, it seems as if the distinction between being required to take responsibility and simply *being* responsible in the first place collapses.

Second, Enoch's claim that his account can accommodate the intuitions that arise in situations of non-culpable harm appears dubious. One way to see this is to consider again Gowans's (1994) scenario involving Craig and the two young women whose car he takes without their permission to save the life of his friend (see 2.2.2 and 4.2.3). On Enoch's view, Craig is morally required to take responsibility for the harm he causes, subsequent to which he faces moral requirements such as apologising, buying the women a gift to express his regret, and so on. But is this plausible? Is it not

more natural to say that Craig faces the requirement to apologise and provide financial compensation regardless of whether he takes responsibility? I think it is. To see why, imagine that Craig refuses to make any acknowledgement to the two women, and a friend of Craig's says, "Craig's behaviour is disgraceful. He really should be apologising and buying them something to show his regret." But then another friend responds, "Oh no, he doesn't need to buy them anything, because he isn't responsible. What he ought to do is to take responsibility, but until he does he doesn't have any other obligations to them." It is likely that we would consider this second friend's response highly peculiar, and that we would agree with the first friend: Craig is subject to certain moral requirements simply in virtue of his non-culpable harm.[178]

Enoch might respond that the above analysis is implausible because it cannot account for the fact that we have the moral intuition that an agent who does not take any steps to acknowledge the harm he has caused (non-culpably) commits a different moral failing from an agent who takes a small step but no subsequent steps. Recall that, in his thought experiment, two drivers non-negligently injure a pedestrian. One of them visits the pedestrian in hospital once and then not again. The other fails to visit the pedestrian at all. On Enoch's view, the moral failing the second driver commits is to fail to take responsibility for the pedestrian's injury. In contrast, the first driver *does* take responsibility for the harm he has caused — he pays a visit to the hospital. His moral failing is in not fulfilling the moral obligations he took on by taking responsibility in the first place. But Enoch's view does not provide a superior explanation of our intuitions here. It is true that the first driver fails differently from how the second driver fails. But, as Goetze notes,

> We can still make this distinction [between the two drivers] … by appreciating the difference in the *extent to which* each driver fails to fulfil his obligations. The driver who never visits at all shirks *all* of his obligations regarding the accident. On the sense of taking responsibility as discharging obligations, he fails to take responsibility at all. By contrast, the driver who visits the child once and never again *partially* fulfils his obligations, but shirks the rest. He begins to take responsibility, in the sense of discharging obligations, but does not follow through…
>
> (Goetze 2021, p. 218)

[178] See Goetze (2021), who makes the point that "what we find fault with … is the agents' failures to *do* the very things that Enoch claims they must first assume obligations to do." (p. 217) Goetze mostly eschews discussion of cases in which an agent non-culpably causes harm, illustrating his point instead with reference to an agent's responsibility for her son's actions, her ancestor's actions, or her country's actions (see table 1, p. 211 for a summary of what he sees as the types of vicarious responsibility). But he does refute an argument of Enoch's involving a non-negligent driver, indicating that he is aware that his account could also be applied to agents who non-culpably cause harm, including agents who face situations of apparent unavoidable moral wrongdoing.

To me, Goetze's explanation is much more plausible than Enoch's, because the second driver is more blameworthy than the first. But if Enoch's view were correct, it is unclear why this would be. Indeed, it seems like the first driver would be less blameworthy than the second. This is because the first driver only fails to act in accordance with one obligation, whereas the second — who on Enoch's view takes responsibility and in doing so accepts a whole host of other obligations — fails to act on many. Of course, Enoch could argue that the obligation to take responsibility was much more morally pressing than each of the obligations that arise from taking responsibility. Perhaps failing to take responsibility is like armed robbery, whereas the obligations that arise from taking responsibility are like petty theft. But this is intuitively highly implausible.

It could be argued that Enoch's other scenario I cite above, involving an agent who is required to take responsibility for his seriously ill friend's daughter in the event of his death, provides better evidence for Enoch's point (let us call this scenario *Ill Father*). If so, perhaps it is the case that *sometimes* there is a duty to take responsibility, and sometimes the agent is simply directly responsible.[179] But it is not clear that Enoch's analysis of *Ill Father* is correct. It is true that this scenario brings out the following two intuitions:

1. the agent does not have an obligation to care for ill father's daughter prior to committing to it; and
2. that, if he were not to commit to it, he would be committing a moral violation.

It therefore seems as if Enoch's notion of a duty to take responsibility is required to reconcile these intuitions. But there is another way of reconciling them, one that does not require us to adopt Enoch's view. As Enoch articulates the scenario, there are "one or two other friends" who are closer to the ill father but who will not (for an unspecified reason) explicitly take on responsibility for his daughter; and it is for this reason that the agent has a duty to take responsibility. But this can also be explained as a collective moral obligation. The ill father's group of friends have a direct collective obligation to care for his daughter; and because the other members of this group have refused to play a role in acting in accordance with this obligation, the moral requirement falls on the agent (as the only other member of the group) to act in accordance with it. There is, then, no need to posit a duty to take responsibility in order to explain our intuitions in *Ill Father*.

[179] Goetze endorses the view that there may well be some instances where an agent has a duty to take responsibility (2021.) And elsewhere, he endorses the view that there is a duty to take responsibility (2022.)

To further illustrate this, I will give another example. Imagine that there is a team of three lifeguards stationed at a popular swimming beach when a strong riptide develops. Most of the bathers manage to effect a successful egress from the water; but one of the lifeguards notices a mother and her child flailing some distance out to sea, getting dragged further and further from the shore with the tide. Plausibly, the team of lifeguards have a collective obligation to render assistance to the woman and her child (let us imagine that the rescue attempt carries some risk, but it is within the range typically considered acceptable.) However, imagine further that two of the lifeguards are new to the job and become paralysed with fear at the prospect of swimming out into the powerful current and attempting a rescue. The final lifeguard is more experienced, and does not become paralysed with fear. Plausibly, in these circumstances, the more experienced lifeguard has an obligation to attempt the rescue. She does not have a duty to take on responsibility for mother and her child, and by virtue of taking responsibility for them acquire an obligation to rescue them. She simply *is* responsible for them — and therefore has particular duties towards them — in virtue of the circumstances. What this shows is that collective obligations can become individual obligations in particular circumstances; and sometimes they can become individual obligations without the individual agent choosing to take them on.[180]

This, I think, explains the fact that it does not seem as if the agent has an obligation to take care of the ill father's daughter, and yet if he were not to promise the ill father that he would do so, he would be committing a moral violation. He does not personally have such an obligation because, in typical circumstances, the obligation lies at a collective level. *Someone* within the group of friends has to take on the duties of caring for the daughter; but no one individual within the group bears the obligation. Typically, there would be discussion within the group; and likely someone would volunteer. Normally, then, someone would choose to take responsibility. However, in *Ill Father,* two members of the group renounce their collective obligation, and so the final member bears the collective obligation alone. The collective obligation becomes, in other words, an individual obligation.[181]

---

[180] It might be objected that collective responsibilities cannot become individual responsibilities in this way, because it is unfair. After all, if others choose to ignore their collective responsibilities, why should the agent who does not suffer as a result? I agree that is in a sense unfair. But this does not change the fact that, where a group of people have an obligation to act, and where the agent knows or has reason to believe that those others who are part of the group which collectively bears the obligation will not act, this does not absolve her of the obligation to act. As Singer famously puts it, "[s]hould I consider that I am less obliged to pull the drowning child out of the pond if on looking around I see other people, no further away than I am, who have also noticed the child but are doing nothing?" (1972, p. 233)

[181] It seems as if it cannot always be the case that, when there is a collective responsibility and some members of the collective fail to acknowledge the responsibility, the entire responsibility devolves upon the individual.

Third, Enoch's responses to the objection that his characterisation of responsibility undermines his view are unconvincing. Let us set aside his first and second responses and examine his third response.[182] Recall that Enoch holds that the duty to apologise is derivative of the duty to take responsibility. If this is correct, then when one is contemplating making the apology, one must not at that moment be required to apologise. Instead, one's ruminations over whether an apology is appropriate must really be cogitations over whether one is required to take responsibility, with the apology being a mere proxy for a more fundamental question. But this does not seem true to our phenomenal experience. It is correct that at times when we are considering whether we ought to do *x,* we ponder whether we ought to perform actions that having a duty to do *x* entails. For instance, imagine one is contemplating whether one has a duty to support an (adult) child as she engages in further university study to retrain as a software engineer. Part of this process of contemplation might involve the consideration of the further obligations that a decision to support the child would generate. However, when we consider whether we ought to apologise (or offer an explanation) for non-culpable harm, we are not pondering the question in order to work out the answer to the antecedent question of whether we ought to take responsibility. Rather, it is the primary question. That is, we are considering whether we are actually responsible in the relevant sense. This, it seems, gives us reason to reject Enoch's account.

Therefore, we have reason to think that an agent who non-culpably causes harm is directly responsible for the harm. There is no intermediate duty to take on responsibility for the harm. Goetze, citing Bernard Williams's classic example of the truck driver, has called this type of responsibility "self-reflexive vicarious responsibility." (2022, p. 1397)[183] However, to say that an agent is

---

For instance, let's say that, prior to 1967, non-Aboriginal Australians had an obligation to give Aboriginal Australians equal rights (including the right to vote, which at the time they lacked.) If every non-Aboriginal Australian but one were determined not to act on this obligation, it seems implausible that the obligation to ensure that Aboriginal Australians got equal rights would fall on the one agent who did not refuse to act on the obligation. However, the difference here is that it is simply not possible for a single agent to effect the changes required to give Aboriginal Australians equal rights. Also, it seems plausible that this one non-Aboriginal Australia would, in such a situation, have obligations above what she would have if a majority of non-Aboriginal Australians were in support of equal rights. For instance, perhaps she would have a greater obligation to try to stir up international outrage against the position of the rest of the Australian population.

[182] I will set aside his first and second responses because even if they are convincing they still leave open the possibility that his view is incoherent in some cases.

[183] In earlier work, Goetze defines self-reflexive vicarious responsibility as "an agent's becoming responsible for their own moral development." (2021, p. 211) This definition seems to exclude situations such as that of Williams's truck driver from being considered as an example of self-reflexive vicarious responsibility. However, he writes later in the article that his view is "preferable as a description of these cases of vicarious responsibility," with the Williams scenario being one that Goetze his referred to in order to illustrate the appeal of his view. So it is clear that Goetze does believe that responsibility that an agent has in cases of non-culpable harm is a type of vicarious responsibility.

vicariously responsible for something is typically to imply that one is responsible in place of another agent. Mellor (2021, p. 241) captures this sense when he defines vicarious responsibility as follows:

> *P1* bears *P2*'s responsibility vicariously if and only if *P1* bears responsibility in the place of *P2*, and *P1*'s responsibility cannot be accounted for without referring to *P2.*

The definition seems to capture exactly why we hold a mother responsible for harm caused by her child, for instance; or why the descendants of slave owners are responsible for the harm caused by the fact that their ancestors owned slaves. In such cases of responsibility, it is clear that there are others in whose place the agent bears responsibility. But in cases of supposed unavoidable moral wrongdoing and other non-culpable harm, it is unclear that there are such other agents. In place of whom does Craig bear responsibility, for instance? In place of whom does Sophie bear responsibility? It appears that in these cases, the answer is in place of nobody. If so, then it is not clear that it is appropriate to describe such agents as being vicariously responsible.

One possible response is to say that agents who cause non-culpable harm bear responsibility in place of the unfortunate set of circumstances that produce the harm. A set of circumstances cannot be held responsible for anything — or if it can, it certainly cannot have any obligations in virtue of its responsibility — and so we hold the agent who was the proximate cause of the harm vicariously responsible.

Whether this is sufficiently convincing to ground the claim that an agent can be self-reflexively vicariously responsible for non-culpable harm is a question I will not pursue here. Instead, I will satisfy myself with the claim that, when an agent causes non-culpable harm, he is directly responsible in a manner that is *akin to paradigmatic cases of vicarious responsibility*. In cases which can uncontroversially be classified as instances of vicarious responsibility, an agent is responsible in virtue of some connection — familial, ancestral, national, religious, etc. — to parties who are morally responsible for the harm. Although in cases of UMW (and other cases of non-culpable harm) there is no agent who is morally responsible for the harm, the agent is nevertheless connected to the harm in such a fashion as to generate responsibility, and this responsibility functions in an identical fashion to paradigmatic cases of vicarious responsibility.

Before moving on, I will deal quickly with a couple of objections. It is doubtless the case that the initial act of apologising or explaining oneself can generate subsequent obligations for the agent. For instance, imagine that when Craig goes to see the young women whose car he took in order to explain himself and take them out for a meal he learns that, because they had to walk a great distance to find help, one of the young women hurt her ankle and has subsequently had to undergo a number

of expensive physiotherapy sessions. In this case, it is plausible that Craig is required to financially compensate the young woman for the physiotherapy. But he would not have faced this requirement without the initial act of apologising. If this is correct, then it seems to fit better with Enoch's view that the initial act of apologising (which constitutes responsibility-taking) generates the subsequent obligations.

But this objection is misguided. It is not that Craig does not have the obligation to compensate the woman for physiotherapy prior to his establishing contact with her. It is that he does not know that he has the obligation. In other words, what this objection is really picking out is an epistemic gap on Craig's part.

Second, it might seem as if this view conflicts with a view I argue for later in this chapter: the view that an agent's obligations are the product of negotiation. It is possible for obligations to arise as a result of interactions between an agent who is directly (non-culpably) responsible and the harmed party. But this does not undermine the point that the notion of direct responsibility better captures the position of an agent who non-culpably commits harm than does the notion of a duty to take responsibility. There are two arguments to show this. First, while not all the moral requirements such an agent has will obtain initially, some of them do obtain from the moment the harm is caused. Second, even if the agent does not have specific requirements prior to engaging in negotiation, he still has obligations — it is just that they are not fully defined. This suggests that such an agent is directly responsible, rather than merely facing a duty to take responsibility, because if the latter were the case none of the moral requirements would obtain except the requirement to take responsibility. And as we have seen, there are good reasons to dismiss this position as implausible.

Therefore, Enoch's elaboration on Wolf's suggestion that there is a duty to take responsibility seems to run into difficulties. More plausible is a view that agents in situations of UMW are directly responsible simply in virtue of standing in a certain relation to some sort of harm or bad outcome.

8.2: Determining One's Obligations

In the previous section, I argued that when an agent is in a situation of moral conflict and, despite making the all-things-considered correct choice nevertheless causes harm, that agent is directly responsible for the harm, rather than merely facing a duty to take responsibility. In this section, I aim to address questions relating to what exactly is required of such an agent. I revise Tamler Sommers's account of blameworthiness — the "negotiating responsibility" view — and contend that the obligations of such an agent are negotiated. I call this view the Negotiating Obligations View. I'll argue

that, in cases of non-culpable harm, the extent of the obligations of an agent who causes non-culpable harm are sensitive to the views of the harmed party, and it is only through a process akin to negotiation that the exact nature of these obligations is determined.

8.2.1: Negotiating Responsibility and Negotiating Obligations

Here, I lay out Sommers's notion of "negotiating responsibility." (2018) Then, I show how it can be adapted to explain what obligations an agent has in the wake of causing harm despite performing the morally right action, all-things-considered.

Sommers articulates his ideas in response to John Doris's arguments on responsibility and agency articulated in *Talking to Our Selves* (2015), and his argument aims to sketch out a view of moral responsibility for a bad act. As what I am interested in in this chapter is non-culpable responsibility — that is, responsibility that does not arise from wrongdoing on the part of the agent — it would seem to be of little assistance to my project. However, as I will show, his ideas can be adapted to provide a promising means of determining what obligations one faces when one is non-culpably responsible.

On Sommers's view, what an agent is morally responsible for, and the extent of this responsibility and what it means for one's obligations, is determined by a process of negotiation or dialogue amongst the affected parties. Sommers holds that philosophical theorising is useless in providing answers to questions about an agent's responsibility; instead, it is up to "the people who are involved in the situation to arrive at — to negotiate — their own responsibility judgments based on what they regard as relevant to the situation…" (2018, no pagination) In order to illustrate this, he uses an example from the film *Manchester by Sea*. In the film,

> Lee [Chandler, the main character] decides to walk to the nearest 24-hour grocery store to get a six-pack of beer. His wife has been sick, and the heating system dries out her sinuses, so their wood fire was the only source of heat. Lee leaves the fire going to warm the house, but forgets to put the screen on the fireplace. A log rolls out. When he returns from the store, the house is on fire. His wife is rescued but their three children burn to death.
>
> (2018, no pagination)

On the view I argued for earlier in this thesis,[184] Lee is at best only very mildly blameworthy here, if he is blameworthy at all. While he may have done some minor wrong in not putting the screen on the fireplace, failing to do so is hardly the apogee of negligence. In the vast majority of cases, such an

[184] That is, a view which rejects resultant moral luck and according to which one's degree of moral responsibility is proportionate to one´s degree of wrongdoing.

action would have no bad consequences, and would warrant (at most) a mild ticking off from his wife. However, Sommers argues that Lee's degree of responsibility cannot be determined through applying theoretical axioms; rather, what matters for his degree of responsibility are "the verdicts of his wife, family, friends, and neighbors." (2018, no pagination) It is how blameworthy the members of this community, especially those closely affected by the consequences of his actions, judge him to be that determines Lee's degree of moral responsibility.

This view is open to a couple of objections. First, Sommers's disavowal of philosophical theorising about responsibility seems misguided, as what he himself is proposing is a philosophical theory of responsibility. It is one that perhaps relies on fewer abstract principles than is typically the case; but it nevertheless is grounded in the intuition that what matters for responsibility is the attitudes of those affected by the morally bad outcome, and that responsibility is determined through some process of negotiation between the affected parties. There are therefore some philosophical constraints on responsibility judgements. It is not as if he is simply holding that anything goes. I think this objection is valid; Sommers's claim that his view is a rejection of philosophical theorising is false. But as I am not committed to the claim that philosophical theorising has no place in determining an agent's degree of responsibility, this is irrelevant as far as my argument goes.

The second problem is more serious for my purposes. It appears that on Sommers's view an agent could be morally responsible to a degree that to an outside observer would appear absurd. Imagine that James is a gay man who lives in a conservative community in rural Texas in the 1990s and forms a relationship with another man. They try to keep the relationship a secret, knowing that it would attract immense opprobrium within their community; however, they are seen together, and their relationship becomes public. James's friends and family are scandalised, and hold James to be extremely morally blameworthy for the relationship and for the harm to family and friends that it causes. On Sommers's view, it seems as if James simply is extremely blameworthy. But such a conclusion is absurd.

Perhaps Sommers has a defence against this charge. After all, on his view responsibility is negotiated. That is, the agent whom others judge to be morally responsible has some say in shaping this responsibility. Perhaps, then, James could respond to his family and friends in order to soften their responsibility judgements. However, unless he is able to negotiate with them to the point where they do not judge him to be blameworthy at all, their responsibility judgements will still be inappropriate. Therefore, Sommers's account needs some additional foundational principle that is not up for negotiation and that prevents an individual or a group from holding an agent to be morally responsible for an action where no moral responsibility is justified.

Fortunately, earlier work by Sommers suggests a solution. In "Partial Desert," (2013) he argues that desert judgements about agents who have done wrong need to consider the attitudes of those affected by the wrongdoing, but should not be determined entirely by these attitudes. On this view, the range within which the punishment for an act of wrongdoing can fall is set by some impartial principle of justice; but where exactly within this acceptable range the appropriate punishment is situated is determined by what the attitudes of the affected parties are towards the wrongdoer and his crime. To illustrate, Sommers asks us to imagine an agent John, who gets drunk after a football game, drives home, and due partially to his inebriated state hits a young girl and kills her. Sommers holds that the acceptable range of punishments in this case is bound by abstract intuitions;[185] John "could not deserve a month long cruise to the Galapagos Islands no matter what the parents wanted. Nor could he deserve to be tortured for twenty consecutive years." (2013, p. 257) However, there is a range of punishments stretching from forgiveness and leniency to a harsh prison sentence that *could* be deserved, and where within this range the agent's desert lies is contingent upon the attitudes of the girl's parents and others affected by the tragedy.[186]

As is clear from earlier parts of this thesis, I do not agree with Sommers that an agent's degree of blameworthiness–and therefore how severely he ought to be punished–can be affected by unintended results of her actions.[187] However, I think that his account *can* successfully be applied to the task of determining what obligations an agent has in the wake of facing a situation of apparent unavoidable moral wrongdoing.[188] Indeed, Sommers himself appears to be aware of this; he writes that it seems as if "the spectrum is broadest in cases where the offence is accidental or the result of negligence." (2013, p. 256) Although this does not capture exactly the plight of agents who face situations of UMW, such agents are in a similar position to those who accidentally cause harm in that the offence does not reflect a bad will. That what an agent owes in the wake of non-culpably causing harm falls within a broad range and is affected by the attitudes of those who are harmed, however, seems exactly right.

---

[185] Here, I mean "acceptable" in the sense of "morally acceptable," not in the sense of what is acceptable under the law.

[186] See Sommers (2016) for more on how the attitudes of harmed parties are relevant to determining the appropriate punishment for the perpetrator.

[187] Interestingly, Sommers seems to think that one's degree of blameworthiness (and therefore the punishment one deserves) is distinct from one's degree of moral responsibility. He holds that the former can be affected by the attitudes of victims, but that the latter cannot: "[j]udgments about John's moral responsibility for the offence are distinct from judgments about what John deserves for having performed it." (2018, no pagination)

[188] It seems plausible that my revised version of Sommers's view can also explain what obligations an agent has when he *has* done wrong. For instance, it appears a fruitful way of explaining what obligations Wolf's (2001) negligent driver has. However, this does not contradict my earlier claim that the view is implausible as a means of explaining an agent's degree of blameworthiness in the wake of wrongdoing.

Here, then, is a version of Sommers's view, revised to account for agential obligations in the wake of non-culpable harm. On my revised version, an agent's degree of blameworthiness cannot be affected by the attitudes of the victims of her wrongdoing. The type of responsibility which can be "negotiated" is non-culpable responsibility. When an agent non-culpably causes harm and thus is responsible, the moral requirements she faces are determined by some abstract principle in concert with the attitudes of the harmed party or parties. What is negotiated, then, is different from what is negotiated on Sommers's view. The negotiation is no longer about what an agent deserves, but about what an agent *owes*, or is obligated to do. As I said earlier, I'll call this the Negotiating Obligations View.

But what of the idea of negotiation? If the obligations are determined by the attitudes of the harmed parties in interaction with abstract intuitions, then this seems like a one-sided process. The perpetrator is absent. There is no negotiation.

This, however, ignores the way in which the harmed party's attitudes are — or ought to be — responsive to the attitudes of the agent who causes the harm. The responsiveness of the agents to one another is what Sommers refers to when he talks about negotiation. I flesh out the nature of this negotiation in the following section.

8.2.2: The Nature of Negotiation

Above, I suggested that the obligations an agent who non-culpably caused harm has are negotiated. I called this the Negotiating Obligations View. However, I have not elaborated on what this negotiation involves. Here, I aim to remedy that.

Let me first articulate the idea in negative terms. Obviously, I do not mean that a formal negotiation takes place, with each person haggling to get the best deal possible for herself. The negotiation, then, is metaphorical, or at least the term is being used somewhat liberally. In order to understand what form this negotiation might take, let us examine what Sommers means by it. Consider the following passage, in which Sommers illustrates his view again with reference to the film *Manchester by Sea*:

> [Responsibility is] a negotiation between the participants, the people affected by the action… . [This means that] [p]erhaps if Lee and his wife had different temperaments, or if they lived in a different town, or if they were more religious, or less religious, they might arrive at a different verdict. Maybe Randi could find a way to forgive Lee, and Lee might even eventually forgive himself. That could be a plausible outcome of negotiation too. Indeed, perhaps the most tragic element of *Manchester by the Sea* isn't that Lee is held morally responsible for an accidental slip. The tragedy is that he is so grief stricken that he *shuts himself off from dialogue*.

> He shuts himself off from negotiation. In a haunting scene eight years after the tragedy, Randi asks Lee to have lunch. She wants to apologize for some of the things she said to him after the incident. But Lee is too broken at this point to continue the dialogue.
>
> (2018, no pagination, emphasis in original)

Sommers sees the negotiation as akin to a "dialogue" in which the harmed party and the agent who has caused the harm each try to understand the perspective of and relevant facts about the other, and to arrive at a responsibility judgement that is acceptable. However, Sommers suggests that it is the harmed party who gets to determine how much blame the agent deserves. For instance, Sommers states that a possible outcome of the negotiation is that "Randi could find a way to forgive Lee," thus limiting Lee's level of blameworthiness.[189] This suggests that it is the harmed party whose judgements about blameworthiness are the ones that ultimately matter.

Note that this does not mean that negotiation is pointless. This is because the harmed party's blame judgements are (or ought to be) responsive to the representations made by the agent who has caused the harm. As Sommers notes, the fact that Lee withdraws from dialogue renders the blame judgements of Randi (and presumably the rest of the community) harsher, as she has nothing to enter into a dialectic with.

On this view, it appears that the harmed party has the ultimate say over the extent of the agent's responsibility (on Sommers's original account) or obligations (on my revised account.) However, this is not entirely the case. It is plausible that, if the harmed party's demands go (and remain) beyond the upper limit of the acceptable range, then she is failing to participate in the negotiation in good faith. The agent is therefore released from considering the views of the harmed party, and is free to respond in a way she considers reasonable or appropriate, using her own wisdom and judgement. As the harmed party is unlikely to be happy with the outcome no matter what the agent does, she will in many cases settle upon the bare minimum. In other cases she may decide to set her obligations higher. The point is that ultimate control of the outcome of the negotiation shifts away from the harmed party, whose failure to negotiate in good faith forfeits her ability to directly influence it. And so it is not the case that the harmed party has total ultimate control over the extent of the obligations.

It might be wondered why a failure on the part of the harmed party to recognise the upper limits of the acceptable range of obligations allows him to determine them. Why does it not result in,

[189] This of course assumes that there is a link between blameworthiness and forgiveness. I'm sceptical of this, but it is easy enough to rework the thought experiment so as to sidestep this issue — perhaps as part of the negotiation, Randi simply comes to understand that he did not intend harm, and so blames him very little.

say, obligations in the middle of the range? Indeed, why does it not create obligations at the top of the range, as this is the closest acceptable set of obligations to those the harmed party has demanded?

The reason is that giving the one who has caused the harm control better respects the demands of fairness. This is so in two ways. First, recall that an agent who has caused non-culpable harm has by definition done nothing wrong — bad luck has put him in the position he is in. In contrast, the victim of the harm *has* done something wrong — she has failed to negotiate in good faith. It seems unfair to penalise the person who is blameless at the expense of the person who is not.[190] Second, by settling upon obligations which exceed the upper limit, the person who has been harmed has shown that, for whatever reason, she lacks good judgement in regard to the matter. It therefore makes sense for the decision-making power to settle upon the other person, who has demonstrated no such lack of judgement.

Sommers's view is an account of moral blameworthiness. As I have said, I think it implausible that blame judgements can depend upon factors outside the agent's control,[191] and so I reject Sommers view as an account of blameworthiness. But the process he articulates *can* account for how obligations are determined in situations of non-culpable harm. Indeed, a revised version of his view — what I have called the Negotiating Obligations View — is highly plausible. In the next section, I offer a justification for this position. For now, I will assume it is correct, and briefly sketch out further the nature of the dialogue.

The form this dialogue takes will vary depending upon the nature of the harm caused. For instance, if I bump into a colleague in the hallway and cause her to spill coffee on her new dress, the dialogue will likely merely consist of an interaction at the site of the incident, in which I assure her I had no ill will, check that she hasn't been burnt by the coffee, and perhaps offer to pay for dry cleaning, depending upon her attitude towards the incident. However, in cases of more serious harm, the dialogue may be more formal in nature. Imagine that a truck driver (through no fault of his own) hits a girl who is playing on the street, badly breaking both her legs. In this case, it is implausible to imagine that a dialogue could occur on the side of the road in the immediate aftermath of the event. The process would likely start there; the truck driver would check on her, express his horror, profusely apologise to her parents, and so on. However, it is unlikely that the dialogue would end

[190] Obviously in cases of particularly severe harm it is understandable for the harmed party's demands to be outside of the acceptable range. If she is upset, it may be difficult for her to properly recognise the acceptable range, or to have the presence of mind to engage in good faith dialogue. In such cases, the person who has caused the harm will, if he exercises wisdom and judgement, likely settle upon obligations near the top of the range.

[191] With some exceptions — see 7.2.2.

there. The situation would likely be too emotionally charged for real progress to be made. In such a situation, the majority of the dialogue would likely take place within a more formal setting. Perhaps the truck driver would contact the parents and arrange a time to call on them at their house, for instance; or maybe he would visit them and the girl at the hospital.[192] Furthermore, it is likely that the dialogue would continue across multiple meetings. Therefore, it is plausible that both the formality and the temporal length of the dialogue are dependent upon the severity of the harm caused. The greater the harm, the greater the formality and the longer the process of negotiating the agent's obligations.

8.2.3: Justifying the Negotiating Obligations View

Above, I have argued that when an agent in a situation of moral conflict is non-culpably responsible for causing harm, the obligations this responsibility generates are arrived at through a process of negotiation. Here, I will show that this view is plausible given the contingency of agential obligation in such situations, and also in light of the justifications I gave earlier (see 5.1) for accepting that an agent who non-culpably causes harm has obligations to the harmed party.

The first justification for the view is an argument Sommers puts forth: that our intuitions are not precise enough to determine the exact extent of an agent's blameworthiness. He argues that, in the absence of a view like his, what is required is (drawing on a phrase from Paul von Hirsch[193]) "a heroic form of intuitionism." (2013, p. 254) What he means by this is that it is not clear that our intuitions regarding desert are precise enough for us to determine in the abstract what punishment is exactly proportionate to a particular form of wrongdoing. He asks, "[w]hat is the deserved punishment for armed robbery? Ten years in prison? Fifteen years? Two years and probation? Flogging?" (2013, p. 254) Our intuitions about desert cannot provide an exact answer to this question. However, intuition *can* give us a rough idea of what punishment is appropriate. For instance, it is obvious that a light wrap on the knuckles is not an appropriate punishment; and nor is being buried alive. Therefore, while intuition may not be heroic enough to pinpoint exactly what an agent who commits a particular crime deserves, it is able to establish a range.

---

[192] Or perhaps the parents of the injured girl would not wish to have any contact with the truck driver, because it triggered thoughts of his truck striking their daughter and the pain she must have experienced and so on. In this case, it would not be that the agent who had caused the harm would owe the parents nothing. Rather, his obligation would be to leave them alone (despite his natural inclination to want to express his sorrow and distress over the incident.)

[193] See Von Hirsch (1992.)

It is plausible that the same is true of our intuitions regarding the obligations of agents who have committed non-culpable harm. What exactly does Craig owe the young women whose car he took without their consent? What does the young man owe his mother when he chooses to go off to fight for his country rather than stay by her side? What does Frankie from *Million Dollar Baby* owe his priest and his church after helping Maggie commit suicide? It appears that there is no objective, precise answer to these questions. Intuition merely suggests rough zones of obligation. For instance, if all Craig does is send a one-line text message to one of the girls saying he hopes they arrived home safely, or if he pushes a brief note under their door, he has not met his obligations. Likewise, if he purchases each of them a new Land Rover and writes a 147-page letter expressing his regret, he has gone wildly beyond what he owes them. But can abstract intuition provide us with an answer to the specific point between these two extremes where Craig's obligations fall? I do not think so. It can only provide us with a broad range. This offers some support for the Negotiating Obligations View.

Second, when we consider a situation in which an agent has caused non-culpable harm, we have an intuition that her obligations are dependent to some extent upon the attitudes of the harmed party. Consider again the situation in which I bump into my colleague in the hallway and cause her to spill coffee on her new dress. Imagine that she is relatively unperturbed about the occurrence, and when I suggest buying her a new dress, she immediately dismisses the idea with a laugh. In this situation, my obligations are minimal — perhaps just to express my regret at the occurrence and to find a cloth to dab away as much of the stain as possible. Now consider a revised version of this scenario. In this version, my colleague is deeply distressed by the incident. Here, it seems as if my obligations are greater than merely to apologise and find a cloth. Depending on the nature of her distress, I may need to pay for dry cleaning, or drop in at her office the next day with flowers to check in on her. If this is correct, it shows that one's obligations are dependent upon the attitudes and judgements of the harmed party. But note that these attitudes and judgements are shaped in part by interactions with the agent who has caused the harm. Let us consider again the scenario in which I bump into my colleague and she is (at first) relatively unperturbed. Were I to simply mumble "sorry," and then rush off, or to express annoyance about the fact that I had collided with her, it is to be expected that her perturbation would grow, and it is likely that her views on the obligations I had towards her would alter. Similarly, in the scenario in which my colleague is deeply distressed and considers me to have greater obligations, my actions could play a role in minimising her sense of my obligations. For instance, if I apologise profusely, empathise with her distress, and so on, it is likely that her views on my obligations would soften.

Importantly for my view, however, we have the intuition that there are limits to what obligations I can have in the above scenario. For instance, if my colleague requests that I buy her a new Volvo, go on a date with her daughter, or purchase her an entirely new wardrobe, it is implausible that I actually have any of these obligations.[194] This suggests that, although the exact extent of my obligations is determined by a process of negotiation or interaction, there is a range within which my obligations can fall which is determined by factors that are independent of the mental states of the parties involved.

Third, the negotiating responsibility account comports with my arguments for an agent having obligations after non-culpably causing harm. Recall that earlier (see 5.1) I argued that there were two plausible justifications for this idea. The first relied directly on intuition. It simply seems that an agent who non-culpably causes harm has moral obligations to the harmed party. When we intuit that *p* we have justification, in the absence of rebutters, for believing that *p*; and as there are no obvious rebutters to the intuition that there are such moral obligations, we are justified in believing that there are indeed such moral obligations. The second justification is consequentialist. On this justification, causal responsibility triggers moral obligations to the harmed party because the consequences of there being such obligations are better than the consequences of there not being such obligations.

This latter justification seems to integrate well with the Negotiating Obligations view. This is because, if an agent who non-culpably causes harm has obligations (partially) for consequentialist reasons, it makes sense that the harm — and how he can ameliorate these harmful consequences — be taken into account in determining how he should act. But the positive consequences of the Negotiating Obligations View go beyond the empowering of the harmed party to negotiate for herself what she is owed and the resultant amelioration of harms. That is, the interactions this view necessitates have positive consequences in another respect. On the view that abstract considerations can determine what obligations the agent who causes the harm has, the agent does not need to interact in a meaningful sense with the harmed party at all. This is because these interactions cannot shape his obligations. And — for the same reasons — nor does the harmed party have to meaningfully interact with the agent who causes the harm, except insofar as such interactions are required in order for the agent to fulfil whatever obligations abstract principles tell him he has. But the Negotiating Obligations View requires such interactions. That these interactions have positive consequences is suggested, I think, in some remarks Nils Christie makes in "Conflicts as Property." In criticising

[194] Indeed, such bizarre demands would likely represent the type of overreach that meant that my obligations were fixed at the lower end of the acceptable range.

Western legal systems which remove the affected parties from personal participation in conflict resolution, Christie comments that the abstract nature of the systems has two detrimental effects: "a loss of pedagogical possibilities," and an increase in "anxiety level and misconceptions." (1977, p. 8) By the first, he means that agents are deprived of opportunities to experience what others value, and therefore to hone their moral sense. By the second, I take him to mean that in the absence of opportunities for interpersonal interaction on the part of the perpetrator of a crime and the victim, the victim retreats into viewing the perpetrator in a stereotypical manner, failing to appreciate her humanity. On Christie's view, then, the interactions between perpetrator and victim have educative value. The Negotiating Obligations View for which I have argued has similar advantages. Through engaging in dialogue in an effort to understand each other, the agent who causes harm and the victim of the harm gain greater insight into each other's motivations, feelings, and values, and therefore emerge from the negotiation better equipped to engage with the world with sensitivity and wisdom.[195]

Before moving on, I will address one objection. At the start of this chapter, I argued against Enoch's view that there is a duty to take responsibility and in favour of a view that the agent is directly responsible. But, it might be argued, this view is in conflict with the negotiating responsibility account of an agent's moral obligations in such a situation. Indeed, the idea that one negotiates one's responsibilities seems to fit better with Enoch's account. This is because, if one's obligations in such a situation are negotiated, the act of entering the negotiation in the first place is required in order to have any obligations. It seems, then, that the act of entering the negotiations is practically equivalent to the act of taking responsibility. An agent who does not enter into negotiations does wrong, but her moral failing is to fail to enter into negotiations, rather than failing to fulfil a whole range of obligations to the harmed party. Therefore, one could argue that if the Negotiating Obligations View is correct then it follows that my arguments in the first part of this chapter are misguided.

But I do not think this is correct. I argued above (8.1.2) that an agent who, in a situation of moral conflict, has non-culpably caused harm need not take responsibility in order to have certain obligations towards the harmed party. This is perfectly consonant with the view that what obligations the agent has are negotiated. To see this, imagine an agent causes non-culpable harm and then simply disclaims all connection to the harm (and so refuses to enter into negotiations.) It is

[195] Jake Wojtowicz makes a similar point about the benefits of agents and harmed parties coming together in the wake of situations of non-culpable harm. Drawing on literature on tort law and restorative justice, Wojtowicz writes that personal interactions mean that "both victim and agent appreciate and properly understand the gravity of the harm and the role it plays in their lives, as well as forms of reparation that help the agent make up for the harm, are practices that might help the victim and agent come to a better understanding of the role of that harm in their lives." (2022, p. 513)

intuitively implausible that the agent's wrongdoing here is merely that he does not enter into negotiations over what his obligations here are. Rather, it is clear that he has obligations to the harmed party prior to any interactions with the harmed party occurring. For instance, if I bump into a colleague in the hallway and cause her to spill her coffee all over her new dress, and then walk off without a word, an observer is unlikely merely to accuse me of failing to engage with the harmed party. She may well accuse me of this, but she is also likely to accuse me of failing to perform some specific actions, such as checking that my colleague had not burnt herself, getting a cloth to help clean up, and so on. This suggests that, while obligations prior to the dialogue beginning may be somewhat vaguely defined, we consider the agent to have some specific obligations. It is not simply that he has an obligation to enter into dialogue and thereby generate obligations. Indeed, if we accept that there is an acceptable range within which one's obligations lie, and the purpose of negotiation is to determine where within this range they lie, then the agent must, prior to negotiation, have obligations. There is no possibility that the result of the negotiation will be that the agent has no obligations at all; at the very least, the agent has the obligations that lie at the lowest end of the range. Therefore, the view that one's obligations are negotiated in such a situation is perfectly compatible with the view that one is directly responsible in cases of non-culpable harm.

8.2.4: Establishing the Acceptable Range

In this section, I examine the question of how we determine the range within which negotiation of the requirements of one's responsibility can take place.

It is plausible that the fundamental principle in determining what an agent's obligations are after causing non-culpable harm is this:

> **Obligation to Repair Harm (ORH):** *When an agent non-culpably causes harm, she has an obligation to repair at least some of the harm.*

This principle is able to account for our intuitions about obligations in cases of supposed unavoidable moral wrongdoing. In cases of grave harm, such as when the truck driver hits and injures a pedestrian, and in more trivial harm, such as when a person bumps into his colleague in the hallway and spills coffee on her, the agent who causes harm plausibly has an obligation to perform at least some actions to repair the damage. But it is implausible that the agent always has the obligation to repair all of the harm. For one thing, it is often not possible. For instance, a truck driver who hits and

kills a pedestrian cannot bring her back from the dead; and nor can a person who knocks over and breaks an inexpensive but sentimentally valuable vase unbreak the vase. Further, an agent cannot be required to do more than to repair the harm, or to perform actions whose relationship to the reparation of harm is unclear. As I noted above, an agent who spills coffee on a colleague cannot be required to buy that colleague three new dresses, and nor can he be required to go on a date with her daughter. In the first case, this is because only one dress has been ruined by coffee, and so at most, the agent owes his colleague one dress. In the second case, dating his colleague's daughter seems utterly unrelated to the harm caused, and so such an obligation is ruled out by the principle ORH.

Of course, in some cases it may be impossible to literally "repair" the harm. For instance, if someone is seriously and irrecoverably injured, the agent who causes the harm cannot effect a repair. However, I am using "repair" here more broadly, to include forms of redress; and in cases such as those involving serious injury, it is redress that is called for.

In summary, an agent who non-culpably causes harm has obligations whose broad range is determined by the principle ORH. Where exactly within this range the agent's obligations fall is specified through a process of negotiation as outlined above.

8.3: Objections to the "Negotiating Obligations" View

In this section, I'll consider three objections to the Negotiating Obligations View I have argued for above. I will begin by considering what I'll call the Objection from Immoral Factors. Then I'll analyse the Evil Agents Objection. Finally, I will look at an objection relating to agent-regret.

8.3.1: Objection from Immoral Factors

One objection to the Negotiating Obligations View is that the harmed party in a situation of unavoidable moral wrongdoing could be influenced in the negotiation process by factors that are irrelevant to the determination of obligations. Such factors include the race or gender of the agent who has caused the harm, or how attractive he or she is, or the agent's social class. Although there is some dispute about the extent of implicit bias, there is evidence to suggest that such bias does distort people's thinking in a range of contexts.[196] But it is problematic to allow such factors to play a role in

[196] See, for instance, Green et. al. (2007) who find that implicit racial bias affects how doctors treat patients; and Pierson et. al. (2020) who find that black drivers are less likely to be pulled over after sunset than before, suggesting that their race plays a role in determining whether police officers pull them over.

determining what obligations an agent has; and this weakens the support for the Negotiating Obligations view.

Let us return to Gowans's scenario involving Craig and the young women to illustrate this. First, imagine that everything is the same in the scenario except that Craig is particularly attractive and charming. When Craig calls on the young women to apologise, his appealing qualities colour their attitudes towards him, and after he has explained his reasons for taking their car without their permission and expressed his deep regret, the young women are happy to absolve him of any further obligations.

Now imagine instead that Craig is not particularly attractive or charming. Indeed, let us imagine that he is rather unattractive and, at least at first, somewhat socially awkward. When he calls on the young women, these qualities unconsciously affect their attitudes towards the harm he has caused them; and as a result the dialogue between them ends in his having greater obligations towards them than he would have had if he had been more appealing in his personal qualities. This seems unfair.

This objection becomes even more powerful when we imagine that race is shaping the harmed party's attitudes towards the obligations of the agent. Let's say that Craig is black, and the two young women are white. Unconsciously, they allow certain stereotypes about black men to creep into their assessment of Craig and therefore to shape their attitudes about what he owes them in the wake of his taking their car without permission. Perhaps, for instance, they unconsciously associate black people with a greater propensity to commit property crime; and so, while on a conscious level they accept that Craig is not blameworthy for his actions, their unconscious bias affects their attitudes towards him and their judgements of his obligations. Because, on the Negotiating Obligations View, the attitudes of the harmed party play a role in determining what obligations an agent has, Craig has greater obligations towards them than if he were white. This appears highly problematic and so could be grounds for rejecting the view.

But this objection can be countered through placing limits on the factors that the harmed party may consider in reflecting on what she is owed. Plausibly, such agents have an obligation to be aware that racial and gender biases could be at play, and to strive to avoid letting factors such as race and gender distort their attitudes. There is evidence to suggest that, if one is aware that one is likely to have certain biases, one can alter one's behaviour to mitigate the impact of these biases. For instance, the findings of one group of researchers suggest that if judges are aware that they may be unconsciously biassed in favour of white people and against black people, they are able to adjust so as to eliminate the impacts of this bias (Rachlinski et. al. 2009). If these findings are accurate, then it

suggests that, if one is committed to combating the effect of implicit bias, one is able to do so effectively.

There are a couple of problems with this response, however. First, it is unlikely that all harmed parties will be committed to eliminating their bias. Indeed, it is plausible that few will. Second, even if agents strive to be aware of the possibility that prejudice is clouding their actions, it is unclear that this will completely negate the bias in every case. As a result, the Negotiating Obligations view appears to entail that certain categories of people will have more obligations than others.

I agree that this is troubling. However, I do not think that it warrants dismissing the Negotiating Obligations view, because the virtues of the account outweigh this drawback. This is in keeping with how we typically deal with problematic bias. For instance, there is evidence to suggest that university professors deemed attractive are regarded as more effective instructors than professors who are not deemed attractive. (Riniolo et. al. 2006) This is obviously unfair; professors are not paid to be models, and students do not attend class for the purposes of erotic titillation. But while we may counsel students to be aware of the potential for attractiveness to skew their perceptions of the effectiveness of their professors, we do not take more drastic steps to eliminate the bias. For instance, we do not force students to attend class blindfolded, or make professors stand behind a curtain to deliver their lectures and seminars. We judge that the value of teachers and students being able to interact in normal human fashion outweighs the benefits such steps would have in eliminating unfairness. The same is true, I claim, of the value of negotiating obligations. It is far from ideal that the obligations of agents will be shaped by factors such as race, gender, and degree of attractiveness. As I suggested above, individuals who have been harmed ought to strive to eliminate the effects of these factors on their attitudes. However, because allowing people to negotiate obligations is of such value, the disvalue of the fact that immoral factors are likely to play some role in shaping the outcomes of these negotiations does not provide a good reason to dismiss the Negotiating Obligations View.

8.3.2: The Evil Victims Objection

In the examples I have used above, the agent who is (non-culpably) harmed is not in the situation due to wrongdoing on his part. However, in some classic examples of apparent unavoidable moral wrongdoing, this is not the case. Indeed, in some cases the agent who is harmed has committed evil acts. Such cases could form the basis for another objection to my view. In these cases, it could be

argued, it is implausible that the agent who takes the correct action all-things-considered, but who in so doing harms the wrongdoer, could be required to negotiate with him over what he owes him. Let us call this the Evil Victims Objection.

Michael Walzer's scenario involving a government official who orders a terrorist to be tortured is an example of a scenario involving an evil agent. Walzer lays out the scenario as follows:

> [A] capital is in the grip of a terrorist campaign, and the first decision the new [political] leader faces is this: he is asked to authorize the torture of a captured rebel leader who knows or probably knows the location of a number of bombs hidden in apartment buildings around the city, set to go off within the next twenty four hours. He orders the man tortured, convinced that he must do so for the sake of the people who might otherwise die in the explosions… .
>
> (1973, p. 166-167)

Let us accept that the politician's use of torture in this case is the morally correct course of action. Even given this, it is plausible that the leader has *some* obligations towards the tortured rebel leader in the wake of ordering him to be tortured;[197] but could it really be the case that his obligations to the rebel leader are dependent upon a dialogue between himself and the rebel? Is this not affording the rebel leader's attitudes and emotions too much respect, given the dastardly and barbaric act of terrorism he was plotting, and that precipitated — and helped to morally justify — the politician's actions?

This line of argument gains even more plausibility if we adjust Walzer's thought experiment somewhat, so that it is similar to the scenario Ivan poses to Aloysha in Dostoevsky's *The Brothers Karamazov*. Imagine that, to ensure the peace and prosperity of his city, the politician, instead of having to torture a rebel leader, has to torture an innocent person.[198] (Perhaps the terrorist group is a collection of psychopaths, for instance, and have said they will detonate a series of devastating blasts unless a child is tortured.) In such a case, it seems as if the child's attitudes and emotions ought to play a far greater role in determining what obligations the politician has towards him in the wake of the torture; and it is highly plausible that this is because the child is completely innocent. He has committed no wrongdoing that has precipitated the situation. But if this is true, then it seems we have found an example which contradicts the Negotiating Obligations View, according to which the range within which the obligations may fall is established by ORH, and then the place within the range is determined according to negotiation.

---

[197] Plausibly, those actually doing the torturing have obligations as well, but for simplicity's sake I will set this issue aside.

[198] Let us assume that such a course of action is all things considered morally correct, given the circumstances.

I acknowledge that this objection requires adjusting the view. But it does not require rejecting it. The Negotiating Obligations View needs to be adjusted to take into account the following idea: if the agent who is harmed in a situation of apparent unavoidable moral wrongdoing committed wrongdoing that has to some degree brought about the situation, then the range within which negotiation can fix the obligations is narrower. That is, the upper boundary is lowered, and lowered in proportion to the extent to which the agent who suffers the harm has done wrong. For instance, in the case of the rebel leader, his wrongdoing is extreme, and so the upper boundary of the range of obligations the politician can have to him is, plausibly, barely above that of the lower boundary. Perhaps an apology on the part of the politician is all that can be required, and all that can be negotiated is its profuseness. There is therefore a very low upper limit on what the politician's obligations can be. But if we imagine a scenario in which the harmed agent has performed only a minor wrong, it seems as if this upper limit is not lowered as much. Take the following situation (although it is not a situation of apparent unavoidable moral wrongdoing, it nevertheless illustrates the point.) A boy is riding an electric scooter and fails to stop at a red light. As a result, he is hit by a car and seriously injured (the light was green for the car, and the driver was not speeding or being negligent in any way, and so the driver is guilty of no wrongdoing.) Even though the non-culpable harm is a result of wrongdoing on the part of the agent who is harmed, plausibly the upper limit on the driver's obligations to the agent is higher than the upper limit in the case of the rebel leader.

Therefore, if an agent has done wrong and is harmed in connection with this wrongdoing, the upper limit on an agent's obligations is correlated with the gravity of any wrongdoing on the part of this agent. This requires an adjustment to the principle ORH that I stated above:

Revised Obligation to Repair Harm:

> *When an agent non-culpably causes harm, she has an obligation to repair at least some of the harm. The extent to which the agent can be obligated to repair the harm is lessened in accordance with the extent to which the harmed party's wrongdoing has played a role in producing the harm.*

Accepting this revised version of ORH defuses the Evil Agents Objection.

8.3.3: Agent-Regret Objection

A related objection has to do with the fact that, in such scenarios as the one from Walzer discussed above, there seems to be a mismatch between the extent of the agent's obligations to the victim and the strength of the agent-regret experienced. If the politician who tortures the rebel leader has only minor obligations towards the rebel leader, then it could be argued that the agent-regret he experiences ought to be quite mild. However, it is not mild; he is tormented by the experience. But if, as I argued earlier (see 5.2.5), the purpose of agent-regret is to alert us to situations in which we have obligations towards others arising from non-culpable harm, then, given his obligations are quite minor, it is unclear why he is tormented.

But I am not committed to the view that the degree of agent-regret one experiences will always track precisely the strength of the agent's obligations. Indeed, there are good reasons to think it will not be. Recall that I endorse Robinson's view of an emotion as a process which begins as a rough, non-cognitive affect and is then refined through cognitive activity (see 1.4.1). In the case of the political leader who tortures the rebel leader, it is to be expected that the initial affect will be extreme moral distress. This is because the rational emotional response to violence *is* moral distress. Typically, when we see that an emotional response is irrational, we work to cultivate the disposition not to experience the emotion. However, in the case of moral distress in situations of apparent unavoidable moral wrongdoing, we have strategic reasons for attempting to cultivate such a disposition — the moral distress works to motivate us to consider what our obligations are in such situations, and to act upon these obligations. But this means that the strength of the emotion will not necessarily correlate with the strength of our obligations. Rather, it will tend to correlate with how morally bad the action is in normal circumstances.

Conclusion

In this chapter, I have clarified the nature of the responsibility that an agent assumes in virtue of causing harm in situations of apparent unavoidable moral wrongdoing. In particular, I have argued against the view — suggested by Susan Wolf and endorsed by David Enoch — that in the wake of non-culpably causing harm an agent has a duty to take responsibility. Instead, I have contended that an agent is simply directly responsible, and that this type of responsibility is at least akin to what is known as vicarious responsibility. This means that, after causing harm despite doing what is all-

things-considered morally right, an agent has certain obligations towards the harmed party, rather than merely an obligation to take on obligations towards the harmed party.

Then, I moved on to analyse the nature of these obligations. Drawing on Sommers's Negotiating Responsibility view, I argued that what exactly these obligations are is determined through a process of negotiation or dialogue between the harmed party and the agent who causes the harm. There are some limits on what the appropriate result of this negotiation can be; these limits are specified through the revised version of the Obligation to Repair Harm (ORH). However, the extent of the agent's obligations within these objective limits is determined through a process of interaction between agent and harmed party. This is justified, I have argued, because it fits with our intuitions about what is required in such cases; and also for consequentialist reasons. One of the reasons for thinking that an agent has any sort of obligation towards someone he non-culpably harms is that it is better overall if agents attempt to repair the harm they have caused; and such harm is more effectively repaired if the agent is responsive to the thoughts and feelings of the harmed party.

Chapter Nine: Concluding Remarks

This thesis has been an investigation of the problem of unavoidable moral wrongdoing. The problem arises because there are situations in which an agent experiences self-directed distress despite having performed the all-things-considered correct action. Such situations present us with a conflict between two compelling intuitions, which I have termed RCW and MDJ. In order to resolve this conflict, we either need to discard one of the two intuitions, or to modify one or both of them such that they can be reconciled.

In this thesis, I have argued that there is a way to convincingly reconcile these intuitions. The key to this reconciliation, I have argued, is Susan Wolf's solution to the problem of resultant moral luck (RML). Wolf denies that there is extant RML. On her view, when an agent is causally responsible for a tragic outcome, but not morally responsible for that outcome — or not entirely so — the negative emotion he experiences is not guilt (or not entirely so) but agent-regret. Furthermore, our sense that the agent has an obligation to compensate the harmed party can be explained by what Wolf calls the "nameless virtue," which requires that an agent take responsibility for outcomes for which he is causally (but not morally) responsible.

Wolf's solution to the problem of RML can, I have argued, be applied to the problem of UMW in order to reconcile the conflict outlined above and to deny that unavoidable moral wrongdoing is possible. The grounds for applying Wolf's view to the problem of UMW is that situations of RML and UMW are similar in three key ways. First, they both appear to involve wrongdoing that is beyond the agent's control, thereby violating the Control Principle. Second, they are phenomenologically similar; in both cases an agent experiences an emotion of distress, and this distress is at least in part self-directed. Third, we have the sense that agents in situations of UMW and RML owe something to the harmed parties. It is therefore plausible that Wolf's solution can successfully resolve the conflict that arises between the intuitions RCW and MDJ in situations of UMW, too.

Applying Wolf's ideas to paradigm cases of UMW buttresses the plausibility of this solution. Take Gowans's scenario involving the hiker Craig and the two young women (see Chapters Two and Four). When Craig takes the young women's car without their permission, he does the morally right thing all-things-considered. However, he is causally responsible for significant harm and so experiences immense agent-regret. This emotion of distress motivates him to act on the nameless virtue and to engage in the process of determining and discharging his obligations to the young women in virtue of the harm he has caused them.

On this account, the emotion of distress that an agent experiences in a situation of UMW is agent-regret. But how is this emotion rational? If the constitutive thought of agent-regret is 'How much better if I'd done otherwise,' and if such a thought is appropriate, then the agent is acknowledging a flaw in his decision procedure. But there is no such flaw; e*x hypothesi,* the agent decides to act in the morally right way, all-things-considered. It is therefore difficult to see how the emotion can be rational.

I've argued that we can overcome this difficulty by recognising that the rationality of agent-regret is strategic rather than representational. Emotions that arise in response to works of fiction provide an apt analogy. If I am watching *King Lear* and am filled with grief as Lear cradles Cordelia's dead body near the end of the play, I do not judge myself to be irrational for experiencing the emotion, despite the fact that neither Lear nor Cordelia are real people, and so cannot suffer or die. The reason for this is not that my emotion is representationally rational — it's not. It's that it is strategically rational. Were I not to experience any emotion as Lear tragically cries out in anguish I would not only deprive myself of pleasure but would also miss out on a valuable lesson about the damaging effects of hubris. The rationality of agent-regret, I've argued, is of a similar nature. The emotion does not accurately represent the world, and so is not representationally rational. However, because of the role it plays in motivating us to fulfil our requirements in the wake of (non-culpably) causing harm, it is of great value to us in navigating a world in which we will inevitably cause such harm. If we are lucky, such harm will be small scale and trifling — a broken promise to meet, say. But on some occasions the harms are far more profound.

What is the relationship between the harm an agent non-culpably causes and his obligations in the wake of the harm? How are these obligations determined? Adapting some work by Tamler Sommers, I've argued that these obligations are set through a process akin to negotiation, and dubbed this the "Negotiating Obligations" view (see Chapter Eight.) An agent cannot determine what he owes a person he has non-culpably harmed without interacting with that person, and understanding the attitude the harmed party takes towards the situation. Imagine a woman who cancels plans for a trip to Thailand with a friend at the last minute because she needs to look after her ill mother. She makes the all-things-considered morally correct choice, but nevertheless causes harm to her friend, who has booked tickets, taken time off work, etc. The woman's obligations here will depend on how that friend is affected by the cancellation. If he is mildly disappointed but brushes it off cheerily and is happy enough to enjoy Thailand on his own, perhaps all that is required is an apology. But if he is despondent at the cancellation — perhaps he finds solo travel unappealing, but cannot get a refund on airfares and hotel bookings at the last minute — then the woman's obligations will be far more

extensive. She may need to compensate her friend for at least part of the cost of the trip. It will depend on how her interactions with her friend proceed.

This is not to say that this process of negotiation can lead anywhere. Rather, there is an acceptable range within which obligations may be negotiated. This range is established through the following principle:

Revised Obligation to Repair Harm:

When an agent non-culpably causes harm, she has an obligation to repair at least some of the harm. The extent to which the agent can be obligated to repair the harm is lessened in accordance with the extent to which the harmed party's wrongdoing has played a role in producing the harm.

This principle accounts for our intuitions about what an agent may be required to do in the wake of causing harm non-culpably. In the example above, for instance, the agent cannot be required to go on a date with her friend's lonely brother, because such an action is unrelated to the harm she has caused. Nor can the agent be obligated to go with her friend on seven vacations to Thailand at a later date, because this would go far beyond repairing the harm. At the upper end of the range, she is perhaps required to compensate him for the cost of the vacation — so he can choose not to go if he wishes — or to partially compensate him, and promise to take a vacation with him at a later date. At the lower end of the range lies a simple apology, and perhaps an offer to get together with him for a meal at her expense. Exactly where between these extremes the woman's obligations lie can only be determined through their interacting with one another with wisdom and sensitivity.

As I have suggested (see Chapter Eight), the Negotiating Obligations View explains an agent's obligations in other situations of non-culpable harm, such as when an agent causes accidental damage — Bernard Williams's famous scenario involving a truck driver who non-culpably hits a pedestrian (1981) is an example. But it is plausible that the view can illuminate an agent's obligations in other situations, too. I'll end by suggesting one such type of situation.

Consider minors who commit horrific acts. In the Australian state of Victoria, when a child between 10 and 14 commits a crime, the prosecution has the burden of showing that the child had the capacity to judge that the act was wrong.[199] In 2023, a 16 year old boy Declan Cutler was stabbed and stomped to death by a group of youths, including a boy who was 13 at the time. A court found

[199] Children under 10 may not be found criminally responsible.

that the prosecution had failed to demonstrate that the 13 year old had this capacity, and adjudged him not guilty.[200] If we assume a retributive theory of punishment, then we have to conclude that the boy — in virtue of his age — did not do wrong.

Declan Cutler's father was distraught at the prospect of the boy walking free, and it is easy to understand why. It seems unsatisfactory to hold that children whose actions produce harm are not responsible in any sense, and therefore have no obligations arising from the harm. The view of non-culpable responsibility, and the Negotiating Obligations View of the obligations that arise from non-culpable harm, could plausibly provide a solution to this. Even if we do not want to say that children below an age threshold are moral agents who can do wrong (and therefore be subject to punishment), we still want to hold them accountable for the harm they cause. If the Negotiating Obligations View can be applied to a situation like the one I outlined above, then even though the 13 year old does not do wrong, he nevertheless still has obligations arising from this conduct.

A case like this is meaningfully different from a situation of apparent UMW in that the child voluntarily causes the harm. Whether this difference means that my view cannot be applied to such situations is a topic for further investigation.

[200] This incident is detailed in a news report by Erin Pearson (2023.)

List of Works Cited

Anscombe, G.E.M. (1958) "Modern Moral Philosophy," *Philosophy,* vol. 33, no. 124, pp. 1-19.

Anderson, E. (1993) *Value in Ethics and Economics*, Cambridge, Harvard University Press.

Aristotle (2012) *Nicomachean Ethics,* Bartlett, R. and Collins, S. (trans), Chicago, University of Chicago Press.

Audi, R. (2004) *The Good in the Right,* Princeton, Princeton University Press.

— (2013) *Moral Perception,* Princeton, Princeton University Press.

Bagnoli, C. (2006) "Breaking Ties: The Significance of Choice in Symmetrical Moral Dilemmas," *Dialectica,* vol. 60, no. 2, pp. 157-169.

Basu, R. (2018) "The Wrongs of Racist Beliefs," *Philosophical Studies* vol. 176, no. 9, pp. 2497-2515.

— (2019) "Radical Moral Encroachment," *Philosophical Issues* vol. 29, no. 1, pp. 9-23.

Blum, C. (2023) "Value Pluralism Versus Value Monism," *Acta Analytica,* vol. 38, no. 4, pp. 627-652.

Brady, M. (2011) "Emotions, Perceptions, and Reasons," in Bagnoli, C. (ed.), *Morality and the Emotions,* Oxford, Oxford University Press, pp. 135-149.

Brink, D. (1994) "Moral Conflict and Its Structure," *The Philosophical Review,* vol. 103, no. 2, pp. 215-247.

Broome, J. (2000) "Incommensurable Values, " in Crisp. R and Hooker, B. (eds) *Well-Being and Morality: Essays in Honour of James Griffin*, New York, Clarendon Press, pp. 21-38.

Buckwalter, W. (2020) "Theoretical Motivations of Ought Implies Can," *Philosophia*, vol. 48, pp. 83-94.

Buckwalter, W. and Turri, J. (2015) "Inability and Obligation in Moral Judgment," *PLoS ONE,* vol. 10, no. 8.

Chituc, V., Henne, P., Sinnot-Armstrong, W., et al. (2016) "Blame, not Ability, Affects Moral 'Ought' Judgments for Impossible Actions: Towards an Empirical Refutation of 'Ought Implies Can,'" *Cognition,* vol. 150, pp. 20-25.

Christie, N. “Conflicts as Property,” *British Journal of Criminology,* vol. 17, no. 1, pp. 15.

Conee, E. (1982) “Against Moral Dilemmas,” *Philosophical Review,* vol. 91, no. 1 pp. 87-97.

Cushman, F. Young, L. and Greene, J. (2010) “Multi-Systems Moral Psychology,” in Doris, J. (ed), *The Moral Psychology Handbook*, Oxford, Oxford University, pp. 47-71.

D’Arms, J. and Jacobson, D. (2000) “The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions, *Philosophical and Phenomenological Research*, vol. 61, no. 1, pp. 65-90.

De Wijze, S. (1996) “The Real Issues Concerning Dirty Hands–A Response to Kai Nielsen”, *South African Journal of Philosophy,* vol. 15, no. 4, pp. 149-151.

— (2024) “Are Dirty Hands Possible?” *The Journal of Ethics,* vol. 28, pp. 187-214.

Domsky, D. (2004) “There Is No Door,” *Journal of Philosophy,* vol. 101, no. 9, pp. 445-464.

Doring, S. (2003) “Explaining Emotion by Action,” *Philosophical Quarterly,* vol. 53, no. 211, pp. 214-230.

Doris, J. (2015) *Talking to Our Selves,* Oxford, Oxford University Press.

Dos Santos, M. (2017) “Walton’s Quasi-Emotions Do Not Go Away,” *Journal of Aesthetics and Art Criticism,* vol. 75, no. 3, pp. 265-274.

Driver, J. (1992) “The Suberogatory,” *Australasian Journal of Philosophy,* vol. 70, no. 3, pp. 286-295.

Dylan, B. (1963) “Masters of War,” *The Freewheelin’ Bob Dylan*, New York, Columbia Records.

Eastwood, Clint (dir.) (2004) *Million Dollar Baby*, Los Angeles, Warner Bros.

Elgin, C. (2008) “Emotion and Understanding,” in Brun, G., Dogluoglu, U., and Kuenzle, D. (eds), *Epistemology and Emotions,* New York, Routledge, pp. 33-49.

Enoch, D. (2012) “Being Responsible, Taking Responsibility, and Penumbral Agency,” in Heuer, U, Lang, G. (eds) *Luck, Value and Commitment: Themes from the Ethics of Bernard Williams*, Oxford, Oxford University Press.

— (2019) “Playing the Hand You’re Dealt: How Moral Luck Is Different from Morally Significant Plain Luck (and Probably Doesn’t Exist), *Midwest Studies in Philosophy*, vol. 18, pp. 257-270.

Enoch, D. and Marmor, A. (2007) “The Case Against Moral Luck,” *Law and Philosophy,* vol. 26, no. 4, pp. 405-436.

Fantl, J and McGrath, M. (2002) "Evidence, Pragmatics, and Justification," *Philosophical Review* vol. 111, no. 1, pp. 67-94.

Fantl, J. (2012) *Knowledge in an Uncertain World*, Oxford, Oxford University Press.

Foot, P. (2002) *Moral Dilemmas: And Other Topics in Moral Philosophy*, Oxford, Oxford University Press.

Fritz, K. (2014) "Responsibility for Wrongdoing without Blameworthiness: How It Makes Sense and How It Doesn't," *Philosophical Quarterly,* vol. 64, pp. 569-589.

Gaita, R. (2004) *Good and Evil: An Absolute Conception*, London: Routledge, 2004.

Gendler, T. (2008) "Alief and Belief," *Journal of Philosophy,* vol. 105, no. 10, pp. 634-663.

— (2010) *Intuition, Imagination, and Philosophical Methodology,* Oxford, Oxford University Press.

Gendler, T. and Kovakovich, K. (2011) "Genuine Rational Fictional Emotions," in Gender, T. (ed) *Intuition, Imagination, and Philosophical Methodology*, Oxford, Oxford University Press, pp. 227-237.

Goetze, T. (2021) "Taking Responsibility and Vicarious Responsibility," *The Monist,* vol. 104, no. 2, pp. 210-223.

— (2022) "Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement," *FAccT22,* pp. 390-400.

Goldie, P. (2000) *The Emotions: A Philosophical Exploration*, Oxford, Clarendon Press.

Gowans, C. (1994) *Innocence Lost: An Examination of Inescapable Moral Wrongdoing*, New York, Oxford University Press.

— (1996) "Moral Theory, Moral Dilemmas, and Moral Responsibility," in Mason, H. (ed) *Moral Dilemmas and Moral Theory,* Oxford, Oxford University Press, pp. 199-215.

Green, A. Carney, D., Pallin, D., Ngo, L., Raymond, K., Iezzoni, L. and Banaji, M., (2007) "Implicit Bias amongst Physicians and its Predictions of Thrombolysis Decisions for Black and White Patients," *Journal of General Internal Medicine,* vol. 22, no. 9, pp. 1231-1238.

Greene, J., Nystrom, L., Engell, A. Darley, J. and Cohen, J. (2004) "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* vol. 44, pp. 389–400.

Greenspan, P. (2000) "Emotional Strategies and Rationality," *Ethics,* vol. 110, no. 3, pp. 469-487.

— (2004) “Emotions, Rationality, and Mind/Body,” in Solomon, R. (ed), *Thinking about Feeling: Contemporary Philosophers on Emotion*, New York, Oxford University Press, pp. 212-229.

Gregory, A. (2017) “The Sorrow and the Shame of the Accidental Killer,” *New Yorker*. Available at: https://www.newyorker.com/magazine/2017/09/18/the-sorrow-and-the-shame-of-the-accidental-killer. Accessed 14 December 2022.

Haidt, J. (2001) “The Emotional Dog and Its Rational Tail,” *Psychological Review,* vol. 104, no. 4, pp. 814-834.

Hales, S. (2015) “A Problem for Moral Luck,” *Philosophical Studies,* vol. 172, no. 9, pp. 2384-2403.

Hanna, N. (2014) “Moral Luck Defended,” *Nous*, vol. 48, no. 4, pp. 683-698.

Hare, R. (1987) “Moral Conflicts,” in Gowans, C. (ed), *Moral Dilemmas*, Oxford, Oxford University Press, pp. 205-239.

Hart, H. and Honore, T. (1985) *Causation in the Law,* 2nd ed., Oxford, Clarendon Press.

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

— (2019a) “Accepting Moral Luck,” in Church, I. and Hartman, R. (eds), *The Routledge Handbook of the Philosophy and Psychology of Luck*, New York, Routledge, pp. 227-238.

— (2019b) “Moral Luck and the Unfairness of Morality,” *Philosophical Studies,* vol. 176, no. 12, pp. 3179-3197.

Holbo, J. (2002) “Moral Dilemmas and the Logic of Obligation,” *American Philosophical Quarterly,* vol. 39, no. 3, pp. 259-274.

Huemer, M. (2005) *Ethical Intuitionism*, Palgrave McMillian, New York.

— (2008) “Revisionary Intuitionism”, *Social Philosophy and Policy*, vol. 25, no. 1, pp. 368-392.

Hurka, T. (1996) “Monism, Pluralism, and Rational Regret,” *Ethics,* vol. 106, no. 3, pp. 555-575.

Hursthouse, R. (1998) *On Virtue Ethics*, Oxford, Oxford University Press.

Jacobson, D. (2012) “Regret, Agency, and Error”, in Shoemaker, D. (ed), *Oxford Studies in Agency and Responsibility Volume 1*, Oxford, Oxford University Press.

Joyce, R. (2000) "Rational Fear of Monsters," *British Journal of Aesthetics,* vol. 40, no. 2, pp. 209-224.

Kamtekar, R., and Nichols, S. (2019) "Agent Regret and Accidental Agency," *Midwest Studies in Philosophy*, vol. 43, no. 1, pp. 181-202.

Kant, I. (2017) *The Metaphysics of Morals*. Edited by L. Denis. Translated by M. Gregor. Cambridge, Cambridge University Press.

Khoury, A. (2018) "The Objects of Moral Responsibility," *Philosophical Studies*, vol. 175, pp. 1357-1381.

Kim, S. (2010) "The Rationality of Emotion Towards Fiction," *Midwest Studies in Philosophy,* vol. 34, no. 1, pp. 106-119.

Krakauer, J. (1996) *Into the Wild,* New York, Anchor Books.

Kumar, V. (2019) "Empirical Vindication of Moral Luck," *Nous*, vol. 53, no. 4, pp. 987-1007.

Kurthy, M. (2017) "Does Ought Imply Can?" *PLoS ONE,* vol. 12, no. 4.

Lane, A. (2025) "Repairing the Damage: Responsibilities, Obligations, and Systemic Racism in Universities, *Problemos,* vol. 107, pp. 163-174.

Latus, A. (2000) "Moral and Epistemic Luck," *Journal of Philosophical Research* 25, 149–172.

Lockhart. J. and Lockhart, T. (2017) "Moral Luck and the Possibility of Agential Disjunctivism," *European Journal of Philosophy,* vol. 26, no.1, pp. 308-332.

Mackenzie, J. (2017) "Agent-Regret and the Social Practice of Moral Luck," *Res Philosophica,* vol. 94, no. 1, pp. 95-117.

Marcus, R. (1980) "Moral Dilemmas and Consistency," *The Journal of Philosophy,* vol. 77 no. 3, pp. 121-136.

Maring, L. (2021) "Beyond Agent-Regret: Another Attitude for Non-Culpable Failure," *Journal of Value Inquiry,* vol. 10, pp. 1-13.

Matravers, D. (2005) "The Challenge of Irrationalism and How Not to Defend it," in Kieran, M. (ed) *Contemporary Debates in Aesthetics and the Philosophy of Art*, Oxford, Blackwell Publishing, pp. 254-264.

McConnell, T. (1978) "Moral Dilemmas and Consistency in Ethics," *Canadian Journal of Philosophy,* vol. 8, no. 2, pp. 269-287.

— (1996) “Moral Residue and Moral Dilemmas,” in Mason, H. (ed), *Moral Dilemmas and Moral Theory*, Oxford, Oxford University Press, pp. 36-47.

McKenna, M. (2012) *Conversation and Responsibility,* New York, Oxford University Press.

Miller, C. (2018) “Generosity: A Preliminary Account of a Surprisingly Neglected Virtue,” *Metaphilosophy,* vol. 49, no. 3, pp. 216-245.

Miller, D. (2005) “Distributing Responsibilities,” *Journal of Political Philosophy,* vol. 9, no. 4, pp. 453-471.

Moen, O. (2016) “An Argument for Intrinsic Value Monism,” *Philosophia,* vol. 44, no. 4, pp. 1375-1385.

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

Morton, J. (2018) “A Dilemma for Streetian Constructivism,” *Southwest* Philosophy Review, vol. 34, no. 1, pp. 133-140.

Muzio, I. (2001) “Emotions and Rationality,” *Croatian Journal of Philosophy,* vol. 1, no. 2, pp. 135-145.

Nagel, T. (1972) “War and Massacre,” *Philosophy and Public Affairs,* vol. 1, no. 2, pp. 123-144.

— (2012). “Moral Luck,” in Sher, G. (ed.), *Ethics: Essential Readings in Moral Theory,* New York, Routledge, pp. 962-976.

Ngozi, C. (2013) *Americanah*, London, 4th Estate.

Nielsen, K. (2000) “There Is No Dilemma of Dirty Hands,” in Shugarman, D. and Rynard, P. (eds), *Cruelty and Deception: The Controversy over Dirty Hands in Politics,* Peterborough, Broadview Press, pp. 139-155.

Nichols, S. and Knobe, J. (2007) “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions,” *Nous,* vol. 41, no. 4, pp. 663-685.

Nelkin, D. (2019) “Thinking Outside the (Traditional) Boxes of Moral Luck,” *Midwest Studies in Philosophy,* vol. 63, pp. 7-23.

Nussbaum, M. (1990) *Love’s Knowledge: Essays on Philosophy and Literature*, New York, Oxford University Press.

— (2000) “The Costs of Tragedy: Some Limits of Cost-Benefit Analysis,” *Journal of Legal Studies,* vol. 29, no. 2, pp. 1005-1036.

— (2011) “Capabilities, Entitlements, Rights: Supplementation and Critique,” *Journal of Human Development and Capabilities,* vol. 12, no. 1, pp. 23-37.

Oshana, M. (1997) “Ascriptions of Responsibility,” *American Philosophical Quarterly*, vol. 34, pp. 71–83.

Payson, J. (2009) “Moral Dilemmas and Collective Responsibilities,” *Essays in Philosophy*, vol. 10, no. pp. 160-182.

Pearson, E. (2023) “‘It’s not Fair’: Father sobs as judge finds 13-year-old not criminally responsible for murder of teen Declan Cutler,” *The Age*, https://www.theage.com.au/national/victoria/judge-finds-13-year-old-not-criminally-responsible-for-murder-of-teen-declan-cutler-20230919-p5e61x.html. Accessed 15 April 2025.

Pierson, E. et al. (2020) “A Large Scale Analysis of Racial Disparities in Police Stops Across the United States,” *Nature Human Behaviour,* vol. 4, pp. 736-745.

Piovarchy, A. (2020) “Blame in the Aftermath of Excused Wrongdoing,” *Public Affairs Quarterly,* vol. 34, no. 2, pp. 142-168.

Rachels, S. (1998) “Counterexamples to the Transitivity of *Better Than*,” *Australasian Journal of Philosophy*, vol. 76, no. 1, pp. 71-83.

Rachlinski, J. Johnson, S., Wistrich, A., and Guthrie, C. (2009) “Does Unconscious Bias Affect Trial Judges?” *Notre Dame Law Review,* vol. 84, no. 3, pp. 1195-1246.

Radford, C. (1975) “How Can We Be Moved by the Fate of Anna Karenina?” *Proceedings of the Aristotelian Society,* vol. 49, pp. 67-80.

— Radford, C. (1995) “Fiction, Pity, Fear and Jealousy,” *Journal of Aesthetics and Art Criticism,* vol. 53, no. 1, pp. 71-75.

Rawls, J. (1999) *A Theory of Justice* (2nd edition), Cambridge, Mass., Harvard University Press.

Richards, N. (1986) “Luck and Desert,” *Mind*, vol. 95, no. 378, pp. 198-209.

Rivera-Lopez, G. (2016) “How to Reject Resultant Moral Luck Alone,” *Journal of Value Inquiry,* vol. 50, pp. 415-423.

Rini, R. (2016) Review of *Moral Failure: On the Impossible Demands of Morality* by L. Tessman, *Mind*, vol. 125, pp. 1227-1236.

Riniolo, T., Johnson, K., Sherman, T., Misso, J. (2006) "Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?" *Journal of General Psychology,* vol. 133, no. 1, pp. 19-35.

Robinson, J. (2004) "Emotion: Biological Fact or Social Construction?", in Solomon, R. (ed), *Thinking about Feeling: Contemporary Philosophers on Emotions,* Oxford, Oxford University Press, pp. 56-83.

— (2005) *Deeper than Reason: Emotion and Its Role in Literature, Music, and Art,* Oxford, Clarendon Press.

Ross, W. (1931/2002) *The Right and the Good,* Oxford, Oxford University Press.

Rosebury, B. (1995) "Moral Responsibility and Moral Luck," *Philosophical Review,* vol. 104, no. 4, pp. 499-524.

Scarre, G. (2017) "The Constitutive Thought of Regret," *International Journal of Philosophical Studies,* vol. 25, no. 5, pp. 569-585.

Schlink, B. (1999) *The Reader*, translated by Janeway, C, London, Vintage.

Schroeder, M. (2010) "Ought, Agents and Actions," *Philosophical Review*, vol. 120, no. 1, pp. 1-41.

Sidgwick, H. (1874/1901) *The Methods of Ethics*, New York, Macmillian.

Singer, P. (1972) "Famine, Affluence, and Morality," *Philosophy and Public Affairs,* vol. 1 no. 3, pp. 229-243.

— (1975) *Animal Liberation*, New York, Harper Collins.

Sinnot-Armstrong, W. (1985) "Moral Dilemmas and Incomparability," *American Philosophical Quarterly,* vol. 22, no. 4, pp. 321-329.

— (1987) "Moral Dilemmas and 'Ought and Ought Not'," *Canadian Journal of Philosophy,* vol. 17, no. 1, pp. 127-139.

— (2006) "Moral Intuitionism Meets Empirical Psychology," in T. Horgan and M. Timmons (eds.) *Metaethics After Moore,* pp. 339-366.

Smith, A. (1759/2005) *Theory of Moral Sentiments*, S. Soares (ed.), Sao Paolo, Metalibri.

Sommers, T. (2013) "Partial Desert," in Shoemaker, D. (ed.), *Oxford Studies in Agency and Responsibility Volume 1*, Oxford, Oxford University Press, pp. 246-262.

— (2016) "The Three Rs: Retribution, Revenge, and Reparation," *Philosophia,* vol. 44, no. 2, pp. 327-342.

— (2018) "Negotiating Responsibility," *The Behavioural and Brain Sciences*, vol. 41, e58.

Statman, D. (2019) "The Definition of Luck and the Problem of Moral Luck," in Church, I. and Hartman, R. (eds), *The Routledge Handbook of the Philosophy and Psychology of Luck,* New York, Routledge, pp. 195-205.

Stocker, M. (1990) *Plural and Conflicting Values*, Oxford, Clarendon.

Stout, R. (2015) "Adopting Roles: Generosity and Presumptuousness," *Royal Institute of Philosophy Supplement,* vol. 71. Pp. 141-161.

Street, S. (2006) "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies,* vol. 127, no. 1, pp. 109-166.

— (2008) "Constructivism about Reasons," in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics Volume 3*, Oxford, Oxford University Press, pp. 207-245.

— (2010) "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* vol. 5, no.5, pp. 363-384.

— (2012) "Coming to Terms with Contingency: Humean Constructivism about Practical Reason," in J. Lenman and Y. Shemmer (eds.), *Constructivism in Practical Philosophy*, Oxford, Oxford University Press, pp. 40-59.

Styron, W. (1979) *Sophie's Choice,* London, Cape.

Sussman, D. (2018) "Is Agent-Regret Rational?" *Ethics,* vol. 128, no. 4, pp. 788-808.

Tannenbaum, J. (2007) "Emotional Expressions of Moral Value," *Philosophical Studies,* vol. 132, pp. 43-57.

— (2018) "Moral Responsibility without Wrongdoing or Blame," in Timmons, M. (ed.) *Oxford Studies in Normative Ethics Volume 7*, Oxford, Oxford University Press, pp. 124-148.

Tappolet, C. (2011) "Values and Emotions: Neo-Sentimentalism's Prospects", in Bagnoli, C. (ed), *Morality and the Emotions,* Oxford, Oxford University Press, pp. 116-134.

— (2016) *Emotions, Value, and Agency,* Oxford, Oxford University Press.

Tempkin, L. (2012) *Rethinking the Good*, Oxford, Oxford University Press.

Teroni, F. (2019) "Emotion, Fiction, and Rationality," *British Journal of Aesthetics,* vol. 59, no. 2, pp. 113-128.

Tessman, L. (2005) *Burdened Virtues: Virtue Ethics for Liberatory Struggles*, New York, Oxford University Press.

— (2015) *Moral Failure: On the Impossible Demands of Morality*, Oxford, Oxford University Press.

— (2016) "*Moral Failure*–Response to Critics," *Feminist Philosophy Quarterly,* vol. 2 no. 1, pp. 1-18.

— (2017) *When Doing the Right Thing Is Impossible*, Oxford, Oxford University Press.

— (2023a) "Moral Injury and Moral Failure," in G. Parsons and E. Wilson (eds.) *How to End a War: Essays on Justice, Peace, and Repair*, Cambridge, Cambridge University Press, pp. 31-58.

— (2023b) "When the Heavens Fall: The Unintelligible and the Unthinkable," *Journal of Ethics,* vol. 27, pp. 495-514.

Vallentyne, P. (1987) "Prohibition Dilemmas and Deontic Logic," *Logique et Analyse*, vol. 30, no. 117-118, pp. 113-122.

Van Zyl, L. (2007) "Can Virtuous People Emerge from Tragic Dilemmas Having Acted Well?" *Journal of Applied Philosophy,* vol. 24, no. 1, pp. 50-61.

Von Hirsch, A. (1992) "Proportionality in the Philosophy of Punishment," *Crime and Justice,* vol. 16, 55-98.

Vranas, P. (2007) "I Ought, Therefore I Can," *Philosophical Studies* vol. 136, no. 2, pp. 167-216.

Wallace, R. (2013) *The View from Here*, Oxford, Oxford University Press.

Walton, K. (1978) "Fearing Fictions," *Journal of Philosophy,* vol. 75, no. 1, pp. 5-27.

— (1990) *Mimesis as Make-Believe*, Cambridge, Harvard University Press.

Walzer, M. (1973) "Political Action: The Problem of Dirty Hands," *Philosophy and Public Affairs,* vol 2, no. 2, pp. 160-180.

Weber, T. (2002) "The Moral Dilemmas Debate, Deontic Logic, and the Impotence of Argument," *Argumentation,* vol. 16, pp. 459-472.

Wedgewood, R. (2009) "The 'Good' and the 'Right' Revisited," *Philosophical Perspectives* vol. 23, no. 1, pp. 499-519.

Whiting, D. (2011) "The Feeling Theory of Emotion and the Object-Directed Emotion," *European Journal of Philosophy,* vol. 19, no. 2, pp. 281-303.

Williams, A. (2011) "Comparing Incommensurables," *Journal of Value Inquiry,* vol. 45, no. 3, pp. 267-277.

Williams, B. (1973) "Ethical Consistency," in *Problems of the Self*, Cambridge, Cambridge University Press.

— (1981) "Moral Luck", in *Moral Luck: Philosophical Papers 1973-1980*, Cambridge, Cambridge University Press.

— (1985/2010) "Morality, the Peculiar Institution," *Ethics and the Limits of Philosophy,* Oxford, Routledge.

Wolf, S. (2001) "The Moral of Moral Luck", *Philosophic Exchange,* vol. 31, no. 1, pp. 1-18.

Wojtowicz, J. (2019) *Agent-Regret in Our Lives.* PhD Thesis. University College London. Available at: https://philpapers.org/archive/WOJAIO.pdf. Accessed 21 January 2025.

— (2022a) "The Purity of Agent-Regret," *Philosophy,* vol. 96, pp. 71-90.

— (2022b) "Agent-Regret, Accidents, and Respect," *Journal of Ethics,* vol. 26, no. 3, pp. 501-516.

Zhao, M. (2020) "Guilt without Perceived Wrongdoing," *Philosophy and Public Affairs,* vol. 48, no 3, pp. 285-314.

Zimmerman, M. (1987) "Luck and Moral Responsibility," *Ethics,* vol. 97, no. 2, pp. 374-386.

— (1996) *The Concept of Moral Obligation,* Cambridge, Cambridge University Press.

— (2002) "Taking Luck Seriously," *The Journal of Philosophy,* vol. 99, no. 11, pp. 553-576.

— (2015) "Moral Luck Reexamined," in Shoemaker, D. (ed), *Oxford Studies in the Philosophy and Psychology of Luck Volume 3,* Oxford, Oxford University Press, pp. 136-159.

— (2019) “Denying Moral Luck,” in Church, I. and Hartman, R. (eds) *The Routledge Handbook of the Philosophy and Psychology of Luck,* New York, Routledge, pp. 216-226.