Moral Worth & Knowledge

Christopher Patrick Benzenberg

B.Phil. Thesis

Word Count:
27.126

19th June 2019



Abstract

Actions can be morally right while lacking moral worth. Kant's prudent shopkeeper, for example, treats his customers fairly only to maximise his long-term profit. Although he's doing the right thing, his actions have no moral worth. It has been suggested that the shopkeeper's actions lack moral worth because they are only accidentally right. *Moral worth excludes luck*. But how should we understand the relevant notion of non-accidental rightness?

To answer this question, this thesis utilises a striking analogy between moral worth and knowledge. Just as actions can be right without having moral worth, beliefs can be true without amounting to knowledge. For instance, guessing the correct outcome of a fair lottery makes for a true belief, but not knowledge. This lack of knowledge has often been explained by the fact that the belief is only accidentally true. *Knowledge excludes luck*.

The notion of epistemic luck has received much attention in the last decades. This thesis draws on the insights from the epistemic literature to clarify the relevant notion of non-accidental rightness in the moral domain. In doing so, two necessary conditions for non-accidental rightness are identified. An action is non-accidentally right only if (i) the action is reliably right; and (ii) the agent has control over the action's rightness.

Table of Contents

1 Introduction **1**

2 The Modal Condition **15**

2.1 Motivation 15

2.2 Tracking 27

2.3 Against Tracking 37

2.4 Safety 47

2.5 Against Safety 53

2.6 Restriction 58

3 The Control Condition **65**

3.1 Motivation 65

3.2 Skepticism and Moral Luck 74

3.3 Weak Control 83

3.4 Restriction 92

4 Conclusion **99**

5 Bibliography **109**

1 Introduction[1]

Actions can be morally right while lacking moral worth.[2] As the first of three examples, consider Kant's famous Shopkeeper case.[3] *S* is a prudent shopkeeper who treats his customers fairly only to maximise his long-term profit. To fill in some details, let us assume that *S* lives in a small village with a fixed customer base and he doesn't overcharge the villagers to maintain a good reputation. There is no doubt that the fair treatment is morally right. However, it seems *S*'s self-interested actions lack moral worth.

Second, imagine the following case of two Brothers.[4] *S* walks past a pond and sees his brother drowning. For reasons reaching back many years, *S* harbours a deep-seated hatred for his brother. In fact, *S* hates his brother so much that he would rejoice to see him drown. However, the brother still owes *S* a lot of money, which will be lost if he drowns. To get his money back, *S* dives into the pond and saves his brother. Although saving the brother was morally right, there is a clear sense in which *S*'s action has no moral worth.

---

1 Throughout this thesis, all references to Kant's writings are to the volume and page of *Kant's gesammelte Schriften* (1900–). Translations are my own.

2 Kant correctly identifies this gap when he suggests that an action can be "in accordance with duty [...] without having genuine moral worth" (4:398). Throughout this thesis, I use the term 'right' to cover mandatory actions.

3 For Kant's original case, see the first section of his *Groundwork* (4:397f.).

4 Isserow discusses a similar case (2018: 3). See also Archer (2016: 241).

Third, and finally, consider the case of a POLITICIAN.[5] *S* is a career politician who is doing politics for money, fame, and power only. However, to win the upcoming election, *S* wants to be portrayed as someone who genuinely cares about the people in his constituency. Following the advice of a PR company, *S* starts serving meals in a soup kitchen once a week. The strategy works, he wins the election. While it's morally right to serve meals in a soup kitchen, we don't think that *S*'s actions have any moral worth.

Before I examine the reasons why *S*'s right actions lack moral worth, I clarify what moral worth is, and what it isn't. In the literature, moral worth is commonly equated with moral praiseworthiness. Arpaly states that "[t]he moral worth of an action is the extent to which the agent deserves moral praise [...] for performing the action" (2003: 69).[6] Indeed, *S*'s actions in the three cases above clearly don't deserve any moral praise. Throughout this thesis, I use the terms 'moral worth' and 'moral praiseworthiness' interchangeably.

It's helpful to contrast the moral worth (or moral praiseworthiness) of an action with four related but different notions. First, we must distinguish the praise*worthiness* of an action from the praise it *actually* receives.[7] Many praiseworthy actions never receive the praise

5 This case is borrowed from Sliwa (2015: 171), (2016: 393).

6 On the relation between moral worth and moral praiseworthiness, see also Sverdlik (2001: 293), Arpaly (2002: 224), (2015: 86), Sorensen (2004: 465n.), (2010: 90), (2014: 295), Markovits (2010: 203), (2012: 309), Archer (2016: 240), Massourd (2016: 691), Sliwa (2016: 395), Isserow (2018: 6), and Das (n.d.).

7 For a further discussion, see Sverdlik (2001: 293), Arpaly (2003: 71), Arpaly & Schroeder (2014: 160f.), Sorensen (2014: 295), as well as Sliwa (2016: 395).

they deserve, and many actions that actually do receive praise might not deserve it. Having a gun pointed at my head, I'll praise even the most despicable deeds. I'm interested in the moral praiseworthiness of actions, and not in whether they have in fact been praised.

Second, many skilled performances are praiseworthy without deserving *moral* praise.[8] Usain Bolt's performance at the 2008 Olympics, for example, deserves to be praised without being morally significant. Christoph Waltz's performance in *Inglorious Bastards* deserves our praise (and an Oscar) without having any apparent moral worth. My interest is in the moral praiseworthiness of actions. However, to avoid lengthy formulations, I hereafter use the term 'praiseworthy' to exclusively denote *moral* praiseworthiness.

Third, the moral worth of an *action* must be distinguished from the moral worth of a *person*. Sliwa notes that "[w]e can evaluate whether actions have moral worth. But we can also morally evaluate agents" (2016: 413).[9] It's easy to see how both evaluations can come apart. On a good day, even the most profligate miscreant can perform a praiseworthy action. And on a bad day, an otherwise virtuous person might perform actions that lack genuine moral worth. I am interested in the moral worth of actions, rather than persons.

---

8 For this distinction, see both Markovits (2010: 203) and Archer (2016: 239f.).

9 For a similar distinction, see Herman (1981: 369), Sverdlik (2001: 296), Arpaly (2002: 231, 240), (2003: 93-96), Markovits (2010: 203, 225), (2012: 301-304), Sorensen (2010: 90n.), Hills (2012: 233), Douglas (2014b), Sliwa (2015: 171), (2016: 413), Archer (2016: 239f.), Massoud (2016: 691), Nelkin (2016: 358), Isserow (2018: 3, 7), as well as Johnson King (n.d.).

Fourth, and finally, there are two valences of moral worth: positive and negative.[10] I equate positive moral worth with praiseworthiness. Negative moral worth, on the other hand, is frequently identified with blameworthiness. Both valences share a common structure. For instance, just as not all right actions deserve praise, it's equally plausible that not all wrong actions deserve blame. My interest is only in the positive moral worth of actions. For the sake of brevity, I hereafter drop the 'positive'-qualification.

Having clarified the notion of moral worth, I now examine the reasons why *S*'s actions in Shopkeeper, Brother, and Politician lack moral worth. Traditionally, it has been argued that an action has moral worth (if and) only if it was motivated in the right way.[11] Sliwa suggests that "[w]hether an action has moral worth depends [...] on the agent's motivation" (2015: 171). On this view, the lack of moral worth in Shopkeeper, Brother, and Politician can be explained by *S*'s morally deficient motivation.

[10] The contrast between moral praiseworthiness and moral blameworthiness has been discussed by Arpaly (2002: 224), (2003: 79-84), (2015: 88), Sorensen (2010), Markovits (2012: 309), Nelkin (2016: 356), as well as Lord (2017: 436).

[11] Broadly speaking, there are two motivational theories of moral worth. First, Kant suggests that an action has moral worth (if and) only if it was done "not from inclination but from duty" (4:399). For an interpretation of Kant's theory of moral worth, see Henson (1979), Herman (1981), de Castro (2005), Guyer (2007), (2018), Timmermann (2009), and Allison (2011: 86-94). Contemporary theories of moral worth state that "our right actions have moral worth only if we perform them for the right reasons" (Markovits 2010: 201). See also Korsgaard (1996a: 60), Stratton-Lake (2000: 60), Nickel (2001: 256f.), Arpaly (2002: 233), (2003: 72), (2015: 86), Hills (2009: 113), (2011: 258f.), (2012: 232-234), (2013: 555), Markovits (2010), McGrath (2011: 133), Sliwa (2012: 191f.), Arpaly & Schroeder (2014), Lord (2015: 37), Archer (2016: 241), Massoud (2016: 695), Whiting (2017: 406), Isserow (2018: 5), and Das (n.d.).

While these motivational considerations are highly attractive, I think that there is another more compelling reason why *S*'s actions in Shopkeeper, Brother, and Politician have no moral worth: *S*'s actions are only accidentally right. If there was more commercial traffic, *S* would have overcharged his customers. If the brother had no debts, *S* would have let him drown. And if there was no election, *S* would not have served meals in the soup kitchen. This line of reasoning draws on the following principle:

Moral Incompatibility

An action φ (performed by a subject *S*) has moral worth only if *S*'s action φ is non-accidentally right.

Moral Incompatibility is universally accepted.[12] According to Kant, "it is not enough that [an action] conforms to the moral law," but its conformity shouldn't be "contingent and precarious" (4:390). Markovits detects the "intuitive thought that in case of morally worthy actions, it's no accident that the agent acts rightly" (2012: 293). Johnson King notes that "all parties [...] to the [...] dispute about moral worth agree that an action lacks moral worth if [...] someone's merely accidentally doing the right thing" (2018a: 3).

[12] Something like Moral Incompatibility has been proposed by Kant (4:390), Herman (1981: 368), Benson (1987: 366-370), Baron (1995: 174), Sverdlik (2001: 303f.), Arpaly (2002: 225), (2003: 71f.), (2015: 101), Hurley (2003: 107), Lackey (2008: 255), Hills (2009), Markovits (2010: 202, 206), (2012: 293f.), Sorensen (2014: 311f.), Douglas (2014a: 81), (2014b: 328-330), Lord (2015), (2017: 439), Sliwa (2015: 171), (2016: 400f.), Archer (2016: 242), Massoud (2016: 695), Isserow (2018: 2), as well as Johnson King (2018b: 99, 101), (n.d.)

I agree. Moral Incompatibility expresses a fundamental truth about moral worth. *Moral worth excludes luck*. However, only a few authors have ever attempted to clarify the relevant notion of non-accidental rightness, and so the truth of Moral Incompatibility is often little more than a truism, a mere platitude. My thesis aims to change this by clarifying the *relevant* notion of non-accidental rightness. For this purpose, it is helpful first to identify three notions of accidental rightness which are, in fact, compatible with moral worth.

First, an action is only accidentally right when the same action could have been wrong. I call this *status luck*. Keeping the moral principle fixed (if there is one), status luck occurs when the right-making features of an action are merely contingent. While feeding peanuts to a starving child might be right in this world, it would be wrong to do so in worlds where the child has a peanut allergy. Status luck is pervasive, but it doesn't undermine moral worth. Giving peanuts to the starving child deserves praise (in this world).

Second, a right action occurs only accidentally when the agent could have lacked the capacities necessary to perform the action. I call this *capacity luck*. For instance, one could have lacked the physical strength to perform a given action. Or maybe, one could have not been alive to perform any actions at all. Like status luck, capacity luck is pervasive but doesn't undermine moral worth. Contingently existing agents can be praised. (On a side note: If ought implies can, capacity luck might be reducible to status luck.)

Third, and finally, a right action occurs only accidentally when the agent could have lacked the evidence for the action's rightness. I call this *evidence luck*. For example, one might be lucky to learn that a particular charity is most effective. However, this doesn't undermine the moral worth of donating to that charity. Evidence luck is compatible with moral worth. (If we think that rightness is evidence-dependent—later I argue against this—, then evidence luck might be reducible to status luck.)

Status luck, capacity luck, and evidence luck specify three notions of accidental rightness which are compatible with moral worth. For these (and further) types of benign luck, Moral Incompatibility turns out to be false. Clearly, however, there is a notion of accidental rightness which is, in fact, incompatible with moral worth. Moral Incompatibility is true for this type of malicious luck. My thesis aims to clarify this *relevant* notion of non-accidental rightness, which makes Moral Incompatibility true.

To this end, I utilise a striking analogy between moral worth and knowledge. We have seen that actions can be right without having moral worth. Likewise, beliefs can be true without amounting to knowledge. For instance, if *S* guesses the correct outcome of a fair Lottery, *S*'s belief is true, but not knowledge. Or suppose that *S* is a Gullible person, believing everything being told. If *S* is told that all whales are mammals, *S* will truly believe that all whales are mammals without gaining knowledge.

There are many potential reasons why *S* lacks knowledge in cases like Lottery and Gullible. In fact, there are as many reasons as there are theories of knowledge. If one thinks that justification is a necessary condition on knowledge, one could argue that *S* lacks knowledge because his beliefs aren't justified. Historically, people might have also argued that *S* has no "adequate evidence" for his beliefs (Chisholm 1957: 16); or that the relevant propositions do not have a "tree of knowledge" (Sosa 1974: 122).

While these considerations might be attractive, I think that there is another more compelling reason why *S* lacks knowledge in Lottery and Gullible. We have seen that right actions lack moral worth if they are only accidentally right. Likewise, true beliefs don't amount to knowledge if they are only accidentally true.[13] Indeed, *S* could have easily guessed a wrong lottery outcome; or *S* could have been told that all whales are fish, rather than mammals. This line of reasoning draws on the following principle:

Epistemic Incompatibility

A belief that *p* (formed by a subject *S*) amounts to knowledge only if *S*'s belief that *p* is non-accidentally true.

[13] Several people have noted the general analogy between epistemic and moral luck. See Nagel (1979: 27), Williams (1993: 242), Latus (2000: 161f.), Axtell (2003: 333), Pritchard (2005: 10, 254), (2006), Ho (forthcoming). Only three have noted the specific analogy between accidental truth and accidental rightness. See Sverdlik (2001: 307), Sliwa (2016: 401), as well as Das (n.d.).

Epistemic Incompatibility is the analogue to Moral Incompatibility. And like Moral Incompatibility, it is (almost) universally accepted.[14] Russell provokingly asks: "Can we ever know anything at all, or do we merely sometimes by good luck believe what is true?" (1912: 204). Unger claims that "a man's belief that *p* is an instance of knowledge only if it is not an accident that the man's belief is true" (1968: 157). And Pritchard identifies, what he takes to be, the "universal intuition that *knowledge excludes luck*" (2005: 1).

While the relevant notion of non-accidental rightness hasn't been properly clarified yet, there has been extensive research on the corresponding notion of non-accidental truth. In this thesis, I draw on the insights from the epistemic literature to clarify the relevant notion of non-accidental rightness in the moral domain. This methodology proves to be extremely fruitful, as we will see. Before moving on, however, we should identify three notions of accidental truth which are, in fact, compatible with knowledge.[15]

---

[14] Defenders of Epistemic Incompatibility include (but don't exhaust) Unger (1967: 172), (1970: 48), Swain (1981: 31), Shope (1983: 192), Lehrer (1990: 12), Engel (1992: 162), (2019), Zagzebski (1994: 72), Latus (2000: 153), Sosa (2000), Williamson (2000: 125), Axtell (2003: 333), Pritchard (2007: 277f.), (2009a: 33), (2009b: 437), (2012: 247), (2014: 594), (2015: 94), Lackey (2006: 284), (2008), Hiller & Neta (2007: 303), Riggs (2007: 330), (2009a: 204), (2014: 627), Kallestrup & Pritchard (2011: 338), Freitag (2014), Gaultier (2014: 490), Schafer (2014: 373), van Riel (2016: 325), Bondy & Pritchard (2018: 3817), Ichikawa & Setup (2018), Faraci (2019: 1), Hirevlä (2019: 1), Beddor & Pavese (forthcoming), Carter et al. (forthcoming), Greco (forthcoming), Greene (forthcoming), and Das (n.d.). To my knowledge, Baumann (2012) and Hetherington (2013) are the only two people to reject Epistemic Incompatibility. For the purpose of this thesis, I simply ignore their unorthodox dissent.

[15] On this topic, see Latus (2000: 165), Lackey (2006: 284), Riggs (2007: 330), Freitag (2014: 44n.), Pritchard (2014: 602n.), as well as van Riel (2016: 325).

First, a belief is only accidentally true when the same belief could have been false. This epistemic version of status luck occurs when the believed proposition is only contingently true. For example, my belief that Trump is the 45th president of the United States is only contingently true because Clinton could have won the 2016 election. While many of our beliefs concern contingent facts, epistemic status luck is compatible with knowledge. There is no question: I know that Trump is the 45th president of the United States.

Second, a true belief occurs only accidentally when the agent could have lacked the capacities necessary to form the belief. For instance, one could have lacked the required sensory capacities to form perceptual beliefs. Or one could have not been alive to form any belief at all. This epistemic version of capacity luck is compatible with knowledge. As Unger puts it: "Though it may be largely accidental that [one] exists or is alive, it may be not at all accidental that [one] is right about various matters of fact" (1968: 160).

Third, and finally, a true belief occurs only accidentally when the agent could have lacked the relevant evidence. This epistemic version of evidence luck is once more compatible with knowledge. Consider the following two examples: If a bank robber accidentally slips his mask, I know who robbed the bank despite my evidence being only lucky (Nozick 1981: 193). Likewise, I can come to know that the Battle of Hastings was fought on 14 October 1066 by using an encyclopaedia which I have luckily won in a lottery.

The epistemic versions of status luck, capacity luck, and evidence luck specify three notions of accidental truth that are compatible with knowledge. Like Moral Incompatibility, Epistemic Incompatibility turns out to be false for these types of benign luck. But again, there is a malicious notion of epistemic luck which is, in fact, incompatible with knowledge. This notion of accidental truth has been the focus of the epistemic literature. I draw on this literature to clarify the corresponding notion of non-accidental rightness.

In this thesis, I use the analogy between moral worth and knowledge to identify two necessary conditions for non-accidental rightness. In the following chapter (§ 2), I examine the modal condition of non-accidental rightness. An action is non-accidentally right only if it is reliably right. In the subsequent chapter (§ 3), I propose the control condition of non-accidental rightness. An action is non-accidentally right only if the agent controls the action's rightness. In the last chapter (§ 4), I summarise and develop the results of this thesis.

However, before I start, I must highlight three essential limitations of my thesis. First, I do not give a full account of non-accidental rightness. While both the modal and the control condition are necessary for non-accidental rightness, neither condition is individually sufficient, nor do both conditions jointly suffice. In fact, I later argue that there are actions which are only accidentally right but satisfy both conditions. In other words, these actions are reliably right, and the agent controls the action's rightness.

Second, even if I were to develop a full account of non-accidental rightness (which I don't), I wouldn't thereby develop a full account of moral worth. There might be actions that are non-accidentally right but lack moral worth. For instance, if one thinks that some condition on the agent's motivation is both necessary for moral worth and not entailed by non-accidental rightness, then non-accidental rightness wouldn't be sufficient for moral worth. Alternatively, one might also think that moral worth is a primitive notion.

Third, and finally, it is commonly noted that moral worth comes in degrees: "[T]he moral worth of actions [...] [is] a matter of degree" (Markovits 2010: 237).[16] Some actions deserve more praise than others. The same applies to non-accidental rightness and truth. The rightness of some actions is more accidental than the rightness of other actions; the truth of some beliefs is more accidental than the truth of other beliefs.[17] However, to reduce complexity, I treat non-accidental rightness and truth as binary notions.

[16] On the degree of moral worth, see also Arpaly (2002: 233-235), (2003: 84), (2015: 87), Markovits (2012: 293f.), Sorensen (2010: 91), Gert (2012: 616f.), Arpaly & Schroeder (2014: 170), Douglas (2014a: 87), (2014b: 325-327), Archer (2016: 239), McNamara (2011: 203), Archer (2016: 239), Massoud (2016: 693), Nelkin (2016: 356), as well as Isserow (2018: 2). Sliwa is the only person denying that the moral worth of actions comes in degrees: "[The] moral praiseworthiness of actions does not admit of degrees" (2016: 414).

[17] More generally, it has been noted that luck comes in degrees. See Williamson (2000: 124), Riggs (2007: 330), Pritchard (2014), (2015), Ichikawa & Setup (2018), Beddor & Pavese (forthcoming), and Stoutenburg (forthcoming).

2 The Modal Condition

In this chapter, I first state and motivate the modal account of non-accidental rightness (§ 2.1). I then specify the modal account along the lines of Nozick's tracking-theory (§ 2.2). While this specification is necessary for non-accidental rightness, it fails to be sufficient (§ 2.3). Next, I argue that the same problem arises when we specify the modal account in terms of safety (§§ 2.4 and 2.5). For these reasons, I finally restrict the modal account to be only a necessary but not sufficient condition for non-accidental rightness (§ 2.6).

2.1 Motivation

Only a few authors have tried to clarify our notion of non-accidental rightness. Those that do analyse non-accidental rightness in modal terms. Sliwa, for example, states that an action is non-accidentally right when "the agent's doing the right thing [is] counterfactually robust" (2015: 178). To avoid awkward formulations, I say that an action is reliably right when its rightness is counterfactually robust. If we think that reliable rightness is both necessary and sufficient for non-accidental rightness, we state a strong modal account:

Strong Moral Reliability

S's action φ is non-accidentally right *if, and only if,*
S's action φ is reliably right.[18]

Note that Strong Moral Reliability does not require that the agent reliably performs a specific *action*. If one reliably performs an action, come what may, the action might only be accidentally right. For instance, if one stubbornly donates to a charity whether or not doing so is right, one's action is only accidentally right. Instead Strong Moral Reliability requires that the action is reliably *right*, i.e., one must only perform the action in some worlds but not others. (I specify the relevant worlds in §§ 2.2 and 2.4.)

To motivate Moral Reliability, it's helpful to first consider its epistemic analogue. Like non-accidental rightness, non-accidental truth is commonly analysed in modal terms. For instance, Pritchard notes: "[T]o eliminate luck [...] from knowledge [...] [we require] a modal principle. [...] [O]ne's true belief should not be false in most near-by possible worlds" (2009a: 33f.). I abbreviate counterfactually robust truth as reliable truth. If non-accidental truth is fully analysed as reliable truth, we again assume a strong modal account:

[18] Something like Strong Moral Reliability has been proposed or indicated by Herman (1981: 363-365), Benson (1987: 370f.), Arpaly (2002: 235), (2003: 86), Douglas (2014a: 87), (2014b: 327), Sorensen (2014: 300, 308-316), Sliwa (2015), (2016: 394), Lord (2017: 439), Benzenberg (n.d.), and Johnson King (n.d.).

Strong Epistemic Reliability

S's belief that *p* is non-accidentally true if, and only if,
S's belief that *p* is reliably true.[19]

Two considerations motivate Strong Epistemic Reliability. First, lucky events in general are often analysed in purely modal terms: "An event is lucky iff it obtains in the actual world but [...] [not] in a wide class of near-by possible worlds" (Pritchard 2007: 279).[20] On this view, an event $\langle e \rangle$ occurs *non*-accidentally if, and only if, the occurrence of $\langle e \rangle$ is reliable (i.e., counterfactually robust). This view is highly intuitive: ⟨winning a lottery⟩ is accidental because ⟨winning a lottery⟩ is by no means reliable.

Strong Epistemic Reliability is an instance of the modal account of non-accidental events in general: ⟨*S*'s belief that *p* being true⟩ occurs non-accidentally if, and only if, the occurrence of ⟨*S*'s belief that *p* being true⟩ is reliable. Insofar as we think that the modal account of non-accidental events in general is intuitive in its own right, we should also think that Strong Epistemic Reliability is intuitive. For surely, the instances of a general account are intuitive when the general account itself is intuitive.

---

[19] I list some proponents of Strong Epistemic Reliability in footnotes 24 and 32.

[20] For a further discussion, see Morillo (1984: 109), Harper (1996), Pritchard (2005: 126-129, 142), (2009a: 33, 37), Cofferman (2007), (2009), and Lackey (2008: 260). Some authors suggest that lucky events must also be significant (Pritchard 2005: 132). For the purpose of this thesis, we can ignore this requirement because beliefs and actions will always be sufficiently significant.

Second, and more importantly, Strong Epistemic Reliability covers a lot of cases. Together with Epistemic Incompatibility, it explains why *S* lacks knowledge in easy cases like LOTTERY, and GULLIBLE. In LOTTERY, *S*'s belief isn't reliably true because he could have easily guessed an incorrect outcome of the lottery. In GULLIBLE, *S*'s belief isn't reliably true because he could have been told, and would have believed, that all whales are fish, not mammals. (For a similar line of reasoning, see page 8.)

The modal account also explains the lack of knowledge in Gettier-cases, like these: After looking at a broken CLOCK, *S* forms the belief that it is 3:00 pm. *S*'s belief turns out to be true—it really happens to be 3:00 pm—, but *S* doesn't gain knowledge (Russell 1948: 154). In the SHEEP case, *S* sees a sheep-shaped object on a field and forms the belief that there is a sheep on the field. The sheep-shaped object is a rock behind which, however, a real sheep is hidden. While *S*'s belief is true, it is not knowledge (Chisholm 1989: 98).

Before I show how Strong Epistemic Reliability (together with Epistemic Incompatibility) covers these Gettier-cases, it is essential to note that theories of knowledge have traditionally struggled to deal with Gettier-cases. The jtb-analysis of knowledge fails because *S*'s beliefs in CLOCK and SHEEP are both true and justified. Moreover, the post-Gettier-debate has shown that most theories of knowledge face Gettier-style counterexamples. For a comprehensive documentation of the post-Gettier-debate, see Shope (1983).

Zagzebski has famously developed "a general rule for the generation of Gettier cases" (1994: 69). Let's call warranted belief any variation on justified belief. As long as there can be warranted false beliefs, we can generate Gettier-cases. For a start, take an unproblematic case of a warranted true belief. Next, make the warranted belief false by adding bad luck. Finally, add a stroke of good luck, making the warranted belief true again. The resulting belief is both "warranted [...] [and] true, but it is not knowledge" (ibid).

In light of Zagzebski's rule, it has been observed that "Gettier cases are often described as cases in which one gets to the truth only as a matter of luck" (Riggs 2009a: 204).[21] People have thus drawn on principles like Epistemic Incompatibility to explain the lack of knowledge in Gettier-cases. In fact, Epistemic Incompatibility gives the right verdict on Clock and Sheep. Looking at a broken clock, *S*'s belief that it is 3:00 pm is only accidentally true. Similarly, *S*'s belief that there is a sheep on the field is true only by accident.

This reasoning essentially relies on Strong Epistemic Reliability, which explains why *S*'s beliefs are only accidentally true. In both Clock and Sheep, *S*'s beliefs are accidentally true because they are not reliably true. *S*'s beliefs could have been false because *S* could

[21] "[K]nowledge must be somehow not dependent on coincidence or luck. This was just the point of the Gettier counterexamples; nothing in the tripart definition excluded knowledge by luck" (Dancy 1985: 134). On this point, see also Sosa (2000: 11), Axtell (2003: 332), Pritchard (2005: 148-152), Hiller & Neta (2007: 303), Freitag (2014: 45), Schafer (2014: 373f.), van Riel (2016: 325), Beddor & Pavese (forthcoming) Greene (forthcoming), as well as Das (n.d.).

have looked at the broken clock at a different time, and there could have been no sheep hidden behind the sheep-shaped rock. That's how Strong Epistemic Reliability (together with Epistemic Incompatibility) explains the lack of knowledge in Gettier-cases.

Drawing on the bipartite motivation for Strong Epistemic Reliability, we can now return to the moral domain and motivate Strong Moral Reliability. First, Strong Moral Reliability is intuitive because it is an instance of the modal account of events in general: The event ⟨*S*'s action *φ* being right⟩ occurs non-accidentally if, and only if, the occurrence of ⟨*S*'s action *φ* being right⟩ is reliable. Strong Moral Reliability obtains its intuitiveness from the modal account of events in general (just like Strong Epistemic Reliability).

Second, and more importantly, Strong Moral Reliability covers a wide variety of cases. In combination with Moral Incompatibility, it explains why *S*'s actions in easy cases such as Shopkeeper, Brother, and Politician lack moral worth. In all three cases, *S*'s actions fail to be reliably right. *S* wouldn't have treated his customers fairly if there was more commercial traffic. *S* wouldn't have saved his brother if he hadn't owed him money. *S* wouldn't have served meals in the soup kitchen if there was no election (see also page 5).

What's more, the modal account can also explain the lack of moral worth in *moral Gettier-cases*. To make this point, I first use Zagzebski's rule to introduce two moral Gettier-cases. Like beliefs, actions can be justified. An action might be justified if the agent has ad-

equate evidence or good reasons to think that the action is right. Any variation on justified action I call warranted action. As long as there can be warranted wrong actions, we can apply Zagzebski's rule. My moral Gettier-cases come in three steps each:

Trolley I: A trolley is rolling towards five people, which are tied to track *A*. If *S* pulls the lever, the trolley is diverted to track *B*. There is no person on track *B*. Being aware of the relevant moral consideration, *S* pulls the lever, thus saving the five people from death. *S* clearly did the right thing (assuming a broadly consequentialist theory of the right). *S* also had good reasons to pull the lever, and so his action was warranted. Finally, given the details provided by this case, it also seems plausible that *S*'s action has moral worth.

Trolley II: As before, a trolley is rolling towards five people on track *A*. If *S* pulls the lever the trolley is diverted to track *B*. This time, however, *S* has bad luck. There are 1000 additional people tied to track *B*. *S* cannot see them because they are located many miles down track *B*. *S* pulls the lever, thereby killing 1000 people (rather than five). Pulling the lever was morally wrong. Yet it seems that, given *S*'s evidence, the action was warranted. However, we clearly wouldn't want to praise *S* for pulling the lever.

Trolley III: Five people are tied to track *A* and 1000 people are tied to track *B*. Like before, *S* cannot see the 1000 people as they are located many miles down track *B*. *S* pulls the lever and the trolley is diverted to track *B*. By a stroke of good luck, a massive rock rolls

on track B, stopping the trolley before it reaches the 1000 people. As a result, it was right for S to pull the lever. What's more, S's action was warranted, given the available evidence. However, I don't think that S's action deserves our praise. It has no relevant moral worth.

Doctor I: S is a doctor who cares about his patients. After a thorough diagnosis, S concludes that one of his patients has disease D_1. Being medically trained, S knows that disease D_1 can only be cured by medication M_1. S prescribes M_1 and the patient recovers within 24h. No doubt, S did the right thing by prescribing medication M_1 (again, assuming a consequentialist theory). Moreover, given S's evidence, S was warranted to prescribe that medication. It also seems that S's action is a plausible candidate for moral worth.

Doctor II: This time, the patient has disease D_2. Due to bad luck, a rare and hitherto unknown medical condition alters the symptoms of disease D_2, making it look as though the patient has disease D_1. After conducting his diagnosis, S thus prescribes medication M_1. However, medication M_1 exacerbates disease D_2. As a consequence, the patient dies within 24h. S was therefore wrong in prescribing medication M_1. Yet, given the evidence, S's action was warranted. However, having killed the patient, S's action deserves no praise.

Doctor III: As before, the patient's medical condition makes it look as though he has disease D_1 when he actually has disease D_2. After conducting his diagnosis, S prescribes medication M_1. The patient takes M_1 and starts to sweat heavily. By a stroke of good

luck, the patient drinks some lemon water. The acid from the lemon chemically reacts with medication M_1 thereby curing the patient. S's prescription turns out to be right. And again, S was warranted in his action. But I don't think S deserves praise for prescribing M_1.

I suggest that Trolley III and Doctor III are moral Gettier-cases. They are instances of warranted right actions without moral worth. At this point, it's important to note that the lack of moral worth doesn't imply that S should be blamed for his actions. S's actions deserve neither praise nor blame. What's more, even if S's actions deserve no praise, it might very well be the case that we would *de facto* praise S for his conduct. We might even think that S deserves to be praised for his moral character (see pages 2-4).

This being said, the lack of moral worth in moral Gettier-cases like Trolley III and Doctor III is difficult to explain. No question: Any account that identifies moral worth with warranted right action fails. But we might even struggle if we think that moral worth requires some condition on the agent's motivation (see page 4). Depending on the details, S's motivation seems morally impeccable. But even if not, any motivational account of moral worth at least lacks a strong grip on moral Gettier-cases like Trolley III and Doctor III.

By contrast, Strong Moral Reliability (together with Moral Incompatibility) has no difficulty explaining the lack of moral worth in Trolley III and Doctor III. Epistemic Gettier-cases are instances of accidentally true belief. Likewise, it seems that moral Gettier-cases

are instances of accidentally right action. We can therefore draw on Moral Incompatibility to explain why S's actions in Trolley III and Doctor III do not deserve to be praised. Pulling the lever and prescribing medication M_1 was right only by accident.

This reasoning essentially relies on Strong Moral Reliability, which explains why S's action's in Trolley III and Doctor III are only accidentally right. In both cases, S's actions are only accidentally right because they are not reliably right. S would have done the wrong thing because the rock could have not rolled onto track B, and the patient might not have drunk the lemon water. That is how Strong Moral Reliability (together with Moral Incompatibility) explains the lack of moral worth in moral Gettier-cases.

Against my discussion of the moral Gettier-cases, one might raise two objections which draw on the distinction between subjective and objective rightness. Put roughly, subjective rightness depends on the agent's "beliefs, knowledge, and evidence" (Lord 2015: 27). Objective rightness depends on features that are independent of the agent's perspective.[22] In utilitarian terms, an action is subjectively right when the evidence suggests that it maximises utility. An action is objectively right when it, in fact, maximises utility.

[22] For a further discussion of subjective and objective rightness, see, among others, Ewing (1948: 112-147), Brandt (1959: 360-67), Russell (1966), Pollock (1979: 109-10), Feldman (1986), (2012), Jackson (1986: 351), Thomson (1986), Oddie & Menzies (1992: 512), Zimmerman (2006), Graham (2010), Hansson (2010), Smith (2010: 65), Parfit (2011: 151-52), Portmore (2011), Dorsey (2012), Carr (2015), Lord (2015: 26f.), Wedgwood (2016), as well as Olsen (2017: 352).

Drawing on this distinction, one might first argue that Strong Moral Reliability (together with Moral Incompatibility) cannot explain the lack of moral worth in Trolley III and Doctor III. In both cases, S's actions are reliably *subjectively* right. For S wouldn't have pulled the lever had he known that there are 1000 people on track B. And S wouldn't have prescribed medication M_1 had he know that the patient has disease D_2. In other words, S would have omitted his actions had they been subjectively wrong.

On top of that, one might argue that moral Gettier-cases are generally impossible. I have previously suggested that an action is warranted (or justified) if the agent's evidence indicates that the action is right. If true, this implies that there can't be actions that are both warranted and *subjectively* wrong. However, without warranted wrong actions, moral Gettier-cases wouldn't get off the ground. The second stage of Zagzebski's rule requires us to introduce bad luck by making the warranted action morally wrong.

Both objections fail because moral worth is about *objective* rightness. S's actions in the second stage (see Trolley II and Doctor II) are (reliably) subjectively right. If moral worth was about subjective rightness, we would thus have to praise S for killing 1000 people on track B or for killing his patient. But this is absurd! Objective rightness avoids both objections: S's actions in Trolley III and Doctor III are not reliably objectively right. And clearly, there can be actions that are both warranted and objectively wrong.

Lastly, it should be noted that I have developed both moral Gettier-cases (Trolley III and Doctor III) in a broadly consequentialist framework. The rightness and wrongness of an action depends (in part) on the action's consequences. I have assumed a consequentialist framework mainly because it is most intuitive and easy to illustrate. However, moral Gettier-cases also arise for other normative theories of the right. They are not restricted to consequentialism. I take Kant's moral theory as an example:

Kant famously claims that we should only act according to maxims "through which [we] can at the same time will that [they] become a universal law" (4:421). Put simply, an action is right if, and only if, its maxim is universalisable. A maxim is universalisable just in case *everyone* could act in accordance with that maxim. The term 'everyone' quantifies over all rational beings and determines the "domain over which maxims need to be universalisable" (Bader n.d.). The universalisability of a maxim depends on the size of the domain.

This interpretation admits the application of Zagzebski's rule. Put very schematically, we start with an unproblematic action whose maxim is universalisable given the domain. Next, due to bad luck, the size of the domain changes. As a consequence, *S*'s maxim isn't universalisable anymore. Finally, by a fluke of good luck, the size of the domain changes once more, making the maxim universalisable again. I contend that the resulting case is a moral Gettier-case. In the context of this thesis, I leave it with these sketchy remarks.

In summary, two considerations motivate Strong Moral Reliability and Strong Epistemic Reliability. Both accounts are instances of our modal understanding of non-accidental events in general. And both accounts cover a wide variety of cases (including Gettier-cases). In the rest of this chapter, I suggest non-accidental rightness in fact requires more than reliable rightness. More specifically, I argue that two specifications of Strong Moral Reliability—tracking and safety—are not sufficient for non-accidental rightness.

2.2 Tracking

In this section, I state, motivate, and refine a specification of Strong Moral Reliability along the lines of Nozick's tracking-theory of knowledge (and non-accidental truth). What exactly does it take for an action to be reliably right? The only person to answer this question was Lord: An action is reliably right if, and only if, one "wouldn't have performed that act if it weren't right" and one "would have performed that act" if it were right (2017: 439). This implies the following specification of Strong Moral Reliability:

Moral Tracking

S's action φ is non-accidentally right if, and only if,
(a) if φ wasn't right, *S* wouldn't perform φ, and
(b) if φ was right, *S* would perform φ.

Lord doesn't realise that Moral Tracking structurally resembles Nozick's tracking-theory of knowledge. In his *Philosophical Explanations* (1981), Nozick proposes the following theory of knowledge: *S* knows that *p* if, and only if, "(1) *p* is true[,] (2) *S* believes that *p*[,] [...] (3) [i]f *p* weren't true, *S* wouldn't believe that *p*[,]" and (4) "if *p* were true, [*S*] would believe [that *p*]" (1981: 172, 176).[23] It appears that conditions (3) and (4) of Nozick's theory are structurally identical to conditions (a) and (b) of Moral Tracking.

To avoid any misunderstandings, Nozick is interested in a theory of knowledge rather than non-accidental truth. However, conditions (3) and (4) of his theory might well be interpreted as specifying our notion of non-accidental truth—at any rate, this is how I read them. In fact, we will see that Nozick's reflections on knowledge transfer to non-accidental truth. In the spirit of Nozick's tracking-theory of knowledge, we can therefore state the following tracking-specification of Strong Epistemic Reliability:

Epistemic Tracking

S's belief that *p* is non-accidentally true if, and only if,
(a) if *p* wasn't true, *S* wouldn't believe that *p*, and
(b) if *p* was true, *S* would believe that *p*.[24]

[23] Nozick states a similar tracking-theory for the moral domain (1981: 317-332). However, he doesn't draw the connection to moral worth and non-accidental rightness. Instead, Nozick is interested in the problem of free will (ibid.).

[24] See also Mackie (1969: 249), Dretske (1970), (1971), Carrier (1971), Lehrer (1971), Goldman (1976), DeRose (1995), Black (2002), Murphy (2005), Roush (2005), Becker (2007), (2008a), (2008b: 356), as well as Black & Murphy (2007).

At this point, it's helpful to introduce some terminology. Condition (a) of Epistemic Tracking is commonly called the *sensitivity* condition. S's belief that p is sensitive (to the truth of p) if, and only if, if p weren't true, S wouldn't believe that p. Condition (b) of Epistemic Tracking is commonly called the *adherence* condition. S's belief that p is adherent (to the truth of p) if, and only if, if p were true, S would believe that p. Lastly, S's belief that p *tracks* (the truth of p) if, and only if, the belief is both sensitive and adherent.

I use the same terminology for the moral domain. Condition (a) of Moral Tracking is the (moral) sensitivity condition. S's action φ is sensitive (to the rightness of φ) if, and only if, if φ wasn't right, S wouldn't have performed φ. Condition (b) of Moral Tracking is the (moral) adherence condition. S's action φ is adherent (to the rightness of φ) if, and only if, if φ was right, S would have performed φ. Lastly, S's action φ tracks (the rightness of φ) if, and only if, the action is both sensitive and adherent.

To move forward, I must clarify the semantics of both conditions in Epistemic Tracking and Moral Tracking. The sensitivity condition is commonly interpreted along the lines of the standard semantics for counterfactuals (Lewis 1973, Stalnaker 1978). On this semantics, S's belief that p is sensitive if, and only if, in all the closest worlds in which p isn't true, S doesn't believe that p (Nozick 1981: 173). Likewise, we might say that S's action φ is sensitive if, and only if, in all the closest worlds in which φ isn't right, S doesn't φ.

The standard semantics make the adherence conditions trivially true. In both the epistemic and the moral domain, the antecedence and the consequence of the adherence conditions are true in the actual world. To avoid this problem, Nozick proposes a different semantics: *S*'s belief that *p* is adherent if, and only if, in most (or all) nearby worlds in which *p* is true, *S* believes that *p* (Nozick 1981: 176). Accordingly, *S*'s action φ is adherent if, and only if, in most (or all) nearby worlds in which φ is right, *S* performs φ.

Having stated both tracking-theories, we must motivate them. It is instructive to first consider Nozick's motivation for Epistemic Tracking. The epistemic sensitivity condition retains the intuitiveness of Strong Epistemic Tracking. Nozick describes it as both "powerful and intuitive" (1981: 173). More importantly, however, the epistemic sensitivity condition covers a wide variety of cases. (I skip Lottery and Gullable.) Nozick introduces the condition "to exclude cases of the sort [...] described by Edmund Gettier" (ibid.).[25]

Indeed, the epistemic safety condition covers both Gettier-cases from the previous section. In Clock, *S* would continue to believe that it is 3:00 pm when in fact it is 3:01 pm because he looks at a broken clock. *S*'s belief thus violates the sensitivity condition and

[25] Many agree with Nozick that the sensitivity condition covers most Gettier-cases. Pritchard writes that "sensitivity-based approaches to knowledge [...] have no difficulty dealing with Gettier-type cases" (2005: 158). Freitag states: "[S]ensitivity accounts draw a lot of support from the [...] fact that they are able to cope with Gettier-type examples, i.e., examples in which the pertinent belief, though justified, is true only by luck" (2014: 45). On this point, see also Sosa (2000: 11), and Pritchard (2009b: 438), (2012: 251-253), (2015: 99f.).

is only accidentally true. Likewise, in SHEEP, *S* would continue to believe that there is a sheep on the field when, in fact, no sheep is hiding behind the rock. Here too, *S*'s belief violates the sensitivity condition and is therefore only accidentally true.

However, the epistemic sensitivity condition is not sufficient for non-accidental truth. In particular, it "does not come into play for necessary truths" (Nozick 1981: 185). Consider the following MATHS case: Upon tossing a coin, *S* forms the true belief that $\sqrt[3]{1728} = 12$.[26] No doubt, *S*'s belief is only accidentally true. Yet it doesn't violate the sensitivity condition because there are no worlds in which $\sqrt[3]{1728} \neq 12$. Depending on one's theory of counterpossibles, the sensitivity condition either doesn't apply or is vacuously satisfied.[27]

To cover cases like MATHS, involving necessary truths, Nozick introduces the epistemic adherence condition (1981: 186). Indeed, the epistemic adherence condition explains why *S*'s belief in MATHS is only accidentally true (and thus not knowledge). There are many nearby worlds in which tossing a coin might lead *S* to form the false belief that $\sqrt[3]{1728} \neq 12$. In other words, if $\sqrt[3]{1728} = 12$ was true, *S* wouldn't have believed it. That's why *S*'s belief violates the moral adherence condition and is only accidentally true.

---

[26] This case is borrowed from Hiller & Neta (2007: 307).

[27] Clarke-Doane, for example, suggests that beliefs in necessary truths "are vacuously sensitive on a standard semantics" (2016: 26). See also Lewis (1986: 114f.), Sosa (1999a: 376), (1999b: 146), Pritchard (2005: 63n.), (2009b), Kripke (2011: 211f.), and Faraci (2019: 10). Nozick further argues that the sensitivity condition fails to cover cases where random chemical processes make the agent form the true belief that he's a brain in a vet (1981: 175f.).

I argue that structurally identical considerations also motivate Moral Tracking. To begin with, the moral sensitivity condition seems to retain the intuitiveness of Strong Moral Reliability. It's intuitive to think that an action is only accidentally right if one would have performed it even if it was false. What's more important, however, moral sensitivity covers a large number of cases. (I skip Shopkeeper, Brother, and Politician.) In particular, it explains why actions in moral Gettier-cases are only accidentally right.

In Trolley III, it would have been (objectively) wrong to pull the lever if the rock hadn't rolled on track B. However, had the rock not rolled on track B, S would still have pulled the lever (see Trolley II). That's why his action violates the moral sensitivity condition and is only accidentally right. In Doctor III, S would have prescribed medication M_1 even if the patient had not drunk the lemon water and died (see Doctor II). Here too, S's action is only accidentally right because it violates the moral sensitivity condition.

Like the epistemic sensitivity condition, the moral sensitivity condition is not sufficient for non-accidental truth. It fails to account for actions that are necessarily right. Consider the following case: S is an Uncle who takes care of his orphan nephew only because he is paid by the government to do so. S's action is only accidentally right. But let's suppose that the orphan nephew has the ineliminable right to be taken care of. In this case, S doesn't violate the moral sensitivity condition because his action cannot be wrong.

Similar to the epistemic domain, cases like UNCLE, involving actions that are necessarily right, can be covered by the moral adherence condition. The moral adherence condition explains why *S*'s action in UNCLE is only accidentally right (and thus lacks moral worth). There are many nearby worlds in which *S* doesn't take care of his orphan nephew because the government doesn't pay him to do so. If it were right to care for the orphan nephew, *S* wouldn't have done it. That's why *S*'s action is not adherent and only accidentally right.

Having stated and motivated both tracking theories, I now discuss an important refinement. Let us start with Nozick's discussion of his GRANDMOTHER case: "A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset" (1981: 179). While the grandmother's belief that her grandson is well is non-accidentally true (and knowledge), it isn't sensitive. For had the grandson been ill, she would have continued to believe that he's well.

This case suggests that the epistemic sensitivity condition is, in fact, not necessary for non-accidental truth. However, rather than rejecting the epistemic sensitivity condition, Nozick proposes to refine Epistemic Tracking: "[W]e must restate our conditions to take explicit account of the ways and methods of arriving at belief" (1981: 179). In other words, we must fix the method *M* via which the agent comes to form his belief. Fixing the method gives us the following refinement of Epistemic Tracking:

Epistemic Tracking Refined

S's belief that p via *M* is non-accidentally true if, and only if,
(a) if *p* wasn't true and *S* were to form a belief whether (or not) *p* via *M*, *S* wouldn't believe that *p* via *M*, and
(b) if *p* was true and *S* were to form a belief whether (or not) *p* via *M*, *S* would believe that *p* via *M*.[28]

Epistemic Tracking Refined gives the right verdict on GRANDMOTHER. If the grandson hadn't been well, and the grandmother would have based her belief on perception (rather than testimony), the grandmother wouldn't have believed that her grandson is well. That's to say, her belief is sensitive (in the refined sense). And clearly, her belief is also adherent (in the refined sense). If her grandson was well and she would have based her belief on perception, she would believe that he is well. The belief is therefore non-accidentally true.

There is a moral analogue to GRANDMOTHER, which makes necessary a refinement of Moral Tracking. In the literature on moral worth, we find the case of an ANIMAL lover: "[*S*] save[s] a group of strangers, at great personal risk" (Markovits 2010: 210). However, had an animal

[28] Several authors fix the method to account for cases such as GRANDMOTHER. See DeRose (1995: 20f.), Becker (2008b: 356f.), Pritchard (2009b: 443f.), (2012), Schafer (2014: 380), van Riel (2016: 325f.), and Das (n.d.). It has further been argued that fixing the method excludes evidence luck. See Pritchard (2005: 156), (2015: 99). The individuation of methods raises two problems which I cannot discuss here. First, there is a debate whether methods should be individuated internally or externally. See Lumper (1984), Shatz (1987), Vogel (1987), Williamson (2000: 156), and Pritchard (2005: 152), (2009b: 444f.). Second, it is unclear how coarse-grained or fine-grained the methods should be individuated. See, for example, Goldman (1976), Nozick (1981: 185), Williamson (2000: 153), Beebe (2004), Becker (2008: 354), and Kripke (2011).

been around, *S* wouldn't have saved the strangers but admired the animal instead. *S*'s rescue clearly deserves praise and is therefore non-accidentally right. But the action is not adherent. In many nearby worlds in which the action is right, *S* doesn't perform it.

Nozick's Grandmother case questions the necessity of the epistemic sensitivity condition. The Animal case questions the necessity of the moral adherence condition. However, in the literature on moral worth, Animal often evokes a much stronger conclusion. Because of Animal, it was argued that we should generally not think about moral worth (and thus non-accidental rightness) in modal terms: "[H]ow the agent might have acted counterfactually is not relevant to the [...] moral worth of actions" (Markovits 2010: 215).[29]

I think this conclusion is premature. Nozick's discussion of Grandmother provides us with a template on how to deal with the Animal case. Rather than rejecting the moral adherence condition or modal reasoning altogether, we should refine Moral Tracking. Nozick proposes to fix the method by which we form a belief. I suggest that the moral analogue to the method of belief is the motive *M* based on which the agent performs the action. Fixing the agent's motive gives us the following refinement of Moral Tracking:

[29] Referring to the Animal case, Isserow notes: "Counterfactual considerations such as these are [...] irrelevant in deciding whether agents are praiseworthy for their actual behaviour" (2018: 3). Johnson King agrees: "[T]he moral worth of an agent's actual action is not hostage to the caprices of her counterparts" (n.d.). For a further discussion of the Animal case and its alleged consequences, see Markovits (2012: 296), Sliwa (2016: 399f.), and Das (n.d.).

Moral Tracking Refined

S's action φ based on *M* is non-accidentally right if, and only if,
(a) if φ wasn't right and *S* were to act based of *M*, *S* wouldn't perform φ based on *M*, and
(b) if φ was right and *S* were to act based on *M*, *S* would perform φ based on *M*.[30]

Moral Tracking Refined covers the Animal case. Let's start with the refined adherence condition. In most nearby worlds in which it is right to rescue the strangers and *S* acts based on his sympathy for other humans (rather than his deep affection for animals), *S* would have rescued the strangers. For this reason, his action is adherent. We can further assume that *S*'s action is also sensitive. *S* wouldn't have rescued the strangers had doing so being wrong (the motive being fixed). As a result, *S*'s action is non-accidentally right.

Note that there are further cases that require additional refinements of the tracking-theories. For example, Nozick discusses "the overdetermination case" where one forms a belief on the basis of more than one method (1981: 182). Nozick's discussion of these cases might also illuminate the moral domain. There's a large debate in the Kant-literature how to handle actions from duty *and* inclination (Yang 2000, de Castro 2005). While interesting, I won't pursue these additional refinements in the context of this thesis.

[30] Citing different reasons, Markovits (2010: 210f.), Arpaly (2015: 99), Sliwa (2016: 399f.), and Isserow (2018: 6) indicate that we should fix the motives. Motives face the same individuations problems as methods (see footnote 28).

In this section, I have stated, motivated, and refined a specification of Strong Moral Reliability along the lines of Nozick's tracking-theory. The result is Moral Tracking Refined, which consists of the moral sensitivity condition and the moral adherence condition. Both conditions specify what it is for an action to be reliably right. In the next section, I argue that Moral Tracking Refined is not sufficient for non-accidental rightness. Some actions are both sensitive and adherent but only accidentally right.

2.3 Against Tracking

Nozick's tracking-theory of knowledge has been subject to various objections.[31] In this section, I focus on an objection raised by Kripke (2011). While directed against Nozick's theory of *knowledge*, Kripke's objection also applies to Epistemic Tracking Refined. The sensitivity and adherence conditions are not sufficient for non-accidental truth. Based on Kripke, I then raise the analogue objection against Moral Tracking Refined. The sensitivity and adherence conditions are not sufficient for non-accidental rightness.

[31] Several people have rejected Nozick's theory of knowledge on the ground that it has skeptical implications. See DeRose (1995: 18), Sosa (1999a: 377), Williamson (2000: 152-156), and Pritchard (2005: 51, 157-160), (2007: 282), (2009b: 441-3). Others criticise that Nozick's theory loses closure under known entailment. See Vogel (1987), Goldman (1983), DeRose (1995: 27), Sosa (1999b: 142f.), (2000: 11), Williamson (2000: ch. 7), Murphy (2005: 332), Pritchard (2005: 39, 47), (2009b: 442f.), Kripke (2011: 190-202), and Engel (2019). For further objections against Nozick, see Forbes (1984), Luper (1987), (2012), Pritchard (2012: 253), (2015: 100), as well as Setiya (2012: 91).

Kripke starts with the famous Barns case: *S* walks through fake-barn-county. In fake-barn-county, there is only one real barn; the other barns are mere facades. The fake barns are indistinguishable from the real barn. While being in fake-barn-county, *S* looks at the only real barn and forms the true belief that there is "a real barn [...] on the field in front of him" (Kripke 2011: 169). Most people think that *S*'s belief doesn't amount to knowledge. Those people will surely agree that *S*'s belief is only accidentally true.

At this stage, Nozick can still explain why *S*'s belief in the Barns case is only accidentally true (and therefore not knowledge). While *S*'s belief satisfies the epistemic adherence condition (for had there been a real barn in front of *S*, he would have believed it), it is not sensitive. *S* wouldn't have changed his belief if there hadn't been a barn on the field in front of him (but a fake barn). This holds true even if we fix the method. *S*'s belief is accidentally true because it violates the sensitivity condition (Nozick 1981: 174f.).

To build his objection, Kripke first distinguishes two scenarios in which *S*'s belief violates the sensitivity condition (2011: 168). In the first scenario, *S*'s belief that there is a barn on the field in front of him would have been false because he stands in front of a fake barn on a different field. In the second scenario, *S*'s belief that there is a barn on the field in front of him would have been false because there is a fake barn on the same field the real barn had been on. In both scenarios, *S* would not abandon his false belief.

Kripke blocks both scenarios by modifying the original BARNS case (2011: 167-169). The modification comes in two parts. First, let us suppose that all fields in fake-barn-county are named in the following series: f_1, f_2, f_3, ... f_n. The real barn was built on field f_3. In Kripke's modification, we don't assess *S*'s belief that there is a barn *on the field in front of him*. Instead, we consider *S*'s belief that there is a barn *on field* f_3. No question: as the old belief, this new belief is only accidentally true (and not knowledge).

This first modification blocks the first scenario in which *S*'s belief isn't sensitive. For if *S* stands in front of a fake barn on a different field, his belief that there is a barn in front of him will be false, but his belief that there is a barn on field f_3 might still be true. This is to say, we don't care about worlds in which *S* stands in front of a fake barn on a different field when we assess whether or not *S*'s belief that there is a barn on field f_3 is sensitive. We only care about those worlds in which there is no barn on field f_3.

Second, let us assume that fake barns can only be erected on a soft ground. Maybe they must be deeply anchored in the ground to withhold strong winds (fill in the details as you like). Real barns, by contrast, can only be erected on rocky grounds. And, in fact, field f_3 has a rocky ground. We can infer, that there wouldn't have been a fake barn on field f_3 if there hadn't been a real barn. Maybe the owner of field f_3 would have erected a silo instead. (Or alternatively, field f_3 would have been left vacant.)

This second modification blocks the second scenario in which S's belief is not sensitive. S's belief that there is a barn on field f_3 is false in those worlds in which there is a silo on field f_3 (rather than a fake barn). Yet it's safe to say that S wouldn't have believed that there is a barn on field f_3 when, in fact, he sees a silo. S's belief is sensitive. However, S's belief is also adherent. If there had been a barn on field f_3, S would have believed it. (In all these counterfactual scenarios, S uses the same method, namely perception.)

I call Kripke's modified barns-case Barns II. We have seen that S's belief in Barns II satisfies both conditions of Epistemic Tracking Refined while only being accidentally true. Barns II thus questions the sufficiency of the sensitivity and adherence conditions for non-accidental truth (let alone knowledge). Note that Barns II is not an isolated counterexample to Epistemic Tracking Refined. We can modify other Gettier-cases to the same effect, i.e., there are many more sensitive and adherent but accidentally true beliefs.

To illustrate this point, I modify Clock and Sheep, starting with Clock II: As before, S looks at a broken clock and forms the true belief that it is 3:00 pm. No doubt, his belief is only accidentally true. But this time a benevolent demon would have altered the clock face had it not been 3:00 pm (Pritchard 2005: 159). S's belief is sensitive and adherent. Had it not been 3:00 pm, S wouldn't have believed it. And had it been 3:00 pm, S would have still believed it. (S uses the same method to form his beliefs: he looks at the broken clock.)

Next, I modify the Sheep case. Sheep II: *S* looks at the sheep-shaped rock and forms the true belief that there is a sheep on the field. *S*'s believe is accidentally true. This time, however, the rock is placed on the field to signal the presence of real sheep (Miracchi 2015). Keeping the method fixed, *S*'s belief is both sensitive and adherent. Hadn't there been a sheep on the field, *S* wouldn't have believed it because there wouldn't have been the rock either. And had there been a sheep on the field, *S* would have still believed it.

Modified Gettier-cases like Barns II, Sheep II, and Clock II are usually cited as counterexamples to Nozick's theory of knowledge (Greco 1999, Lackey 2008, Schafer 2014). I have argued that they are also counterexamples to the tracking theory of non-accidental truth. I contend that there are many more cases in which the agent's belief satisfies both conditions of Epistemic Tracking Refined while only being accidentally true. We can therefore confidently conclude that sensitivity and adherence are not sufficient for non-accidental truth.

In the second half of this section, I employ the epistemic counterexamples as a template to devise analogue counterexamples against Moral Tracking Refined. More specifically, I first develop the moral analogue to Kripke's Barns II case. I then argue that this case is not an isolated counterexample by modifying both moral Gettier-cases from the previous sections (Trolley III and Doctor III). These cases demonstrate that the moral sensitivity and adherence conditions are not sufficient for non-accidental rightness.

The moral analogue to the original BARNS case is the following BEGGARS case: *S* walks through fake-beggar-county. In fake-beggar-county, there is only one real beggar. The other beggars are all fake, meaning they are not indigent. What's more, the fake beggars are virtually indistinguishable from the real beggar. *S* makes a donation to the person on the slot in front of him. As it happens, *S* stands in front of the only real beggar. His action is right. However, it seems that *S*'s donation is only (objectively) right by accident.

At this stage of my objection, Moral Tracking Refined still explains why *S*'s donation is only accidentally right. To be sure, *S*'s action satisfies the moral adherence condition. *S* would have donated to the person on the slot in front of him if doing so had been right. However, *S*'s action is not sensitive. *S* wouldn't have omitted his action hadn't there been a beggar on the slot in front of him (but a fake beggar). Moreover, *S*'s donation to the fake beggar would have been based on the same motive, namely compassion.

As in the epistemic domain, we must distinguish two scenarios in which *S*'s action is not sensitive. In the first scenario, *S*'s donation to the person on the slot in front of him would have been wrong because *S* stands in front of a fake beggar on a different slot. In the second scenario, *S*'s donation to the person on the slot in front of him would have been wrong because there would have been a fake beggar on the same slot the real beggar had been on before. In both scenarios, *S* would not have omitted the wrong donation.

Kripke teaches us how to block both scenarios. We must make two modifications to the BEGGARS case. First, we assume that all slots in fake-beggar-county are named in the following series: s_1, s_2, s_3, ... s_n. The real beggar sits on slot s_3. We can now fix a different description of *S*'s action. We no longer asses the description *donating-money-to-the-person-on-the-slot-in-front-of-him*. Instead we asses *S*'s action under the description *donating-money-to-the-person-on-slot-s_3*. Even under this new description, *S*'s action is only accidentally right.

This first modification blocks the first scenario in which *S*'s action isn't sensitive. If *S* stands in front of a fake beggar on a different slot, his donation to the person on the slot in front of him will be wrong, but the donation to the person on slot s_3 might still be right. To assess whether *S*'s donation to the person on slot s_3 is sensitive, we don't care about worlds in which *S* donates money to the person on the slot in front of him. We only care about those worlds in which it is wrong to donate money to the person on slot s_3.

Second, let us assume that fake-beggar-county has a law which ensures that only real beggars can sit on slot s_3. This law is rigorously enforced. In order to sit on slot s_3, people have to prove that they are indigent. And if no real beggar sits on slot s_3, a police officer blocks the slot to ensure that no fake beggar is sitting down. From this, it follows that there wouldn't have been a fake beggar on slot s_3, hadn't there been the real beggar on the slot. Instead, slot s_3 would have been blocked by a police officer.

This second modification blocks the second scenario in which S's action isn't sensitive. S's donation to the person on slot s_3 is wrong in those worlds in which the slot is blocked by a police officer (rather than a fake beggar). But surely, S wouldn't have donated money to the police officer. S's action is therefore sensitive. But S's action is also adherent. If there had been a real beggar on slot s_3, S would have made the donation (or so we can assume). And in all these scenarios, S acts on the same motive: compassion.

I call the modified beggars-case Beggars II. I have argued that S's action in Beggars II satisfies both conditions of Moral Tracking Refined while only being accidentally right. This argument raises substantial doubts about whether the moral sensitivity and adherence conditions are sufficient for non-accidental rightness. These doubts are reinforced by the fact that we can modify other moral Gettier-cases to the same effect. Beggars II is not an isolated counterexample to Moral Tracking Refined.

I illustrate this point by modifying the original moral Gettier-cases Trolley III and Doctor III. Consider Trolley IV: Again, there are five people on track A and (out of S's sight) 1000 people on track B. S pulls the lever to divert the trolley to track B. Before the trolley reaches the 1000 people, the rock rolls on track B. However, had the rock not rolled on the track, Superman would have protected all people *out of S's sight*. This modification doesn't change the fact that S's action is (objectively) right only by accident.

However, S's action in Trolley IV is both sensitive and adherent. Let's start with the sensitivity condition. S wouldn't have pulled the lever if doing so had been wrong for him. The closest worlds in which it's wrong for S to pull the lever are those in which there are more than five people tied to track B *that are in S's sight*. We assume that all people that are out of S's sight are protected by Superman and therefore not in danger. S wouldn't have pulled the lever had he seen more than five people on track B.

The adherence condition is next. S would have pulled the lever if doing so had been right. Here is why: The nearby worlds in which it is right to pull the lever are those worlds in which either the rock or Superman stop the trolley before it reaches the 1000 people. But S would have pulled the lever in all those worlds because he is only aware of the five people on track A, not the 1000 people on track B. On top of all this, we can plausibly assume that S always acts on the same rational motive (e.g., minimising the number of deaths).

Doctor IV makes two modifications to Doctor III: First, let us assume that all water (not just lemon water) is acidic. Remember that the acid chemically reacts with medication M_1 thereby curing the patient with disease D_2. Second, let us assume that medication M_1 immediately kills every healthy person—the acid doesn't help. In this modified case, S again prescribes medication M_1 to cure the patient whom he mistakenly believes to have disease D_1. Despite the modifications, S's action is still only accidentally right.

However, S's action in Doctor IV is both sensitive and adherent. I start with the sensitivity condition: Had it been wrong to prescribe medication M_1, S wouldn't have done so. For a start, note that S's action would not have been wrong in those worlds in which the patient has disease D_2. That's because the patient reliably consumes the acid. The closest worlds in which the prescription is wrong, are those in which the patient is healthy (he would immediately die). But surely, S doesn't prescribe M_1 to a healthy person.

S's action also satisfies the adherence condition. Had it been right to prescribe medication M_1, S would have done so. Here is why: The nearby worlds in which the prescription is right are those worlds in which the patient has disease D_2 and drinks the acidic water. S would have prescribed medication M_1 in all those worlds because the patient's rare medical conditions make it look as though he has disease D_1, which requires medication M_1. Moreover, S always acts on the same motive, namely his care for the patients.

In sum, modified moral Gettier-cases like Beggars II, Trolley IV, and Doctor IV are counterexamples to Moral Tracking Refined. They demonstrate that sensitivity and adherence are not sufficient for non-accidental rightness. However, the problem raised by these modified moral Gettier-cases is not restricted to the tracking theory of non-accidental rightness. In the following sections, I argue that Beggars II, Trolley IV, and Doctor IV are also counterexamples to a specification of Strong Moral Reliability in terms of safety.

2.4 Safety

In this section, I state, motivate, and refine a specification of Strong Moral Reliability in terms of safety. For this purpose, I again use the resources provided by epistemology. Many authors reject Nozick's tracking-theory (especially the sensitivity condition) in favour of a safety-based account.[32] But what is safety? Here is Sosa's definition: "[A] belief is safe iff *S* would not have held it without it being true" (2000: 14). Put differently, *S*'s belief that *p* is safe if, and only if, if *S* were to believe that *p*, *p* would be true.

Many authors take safety as a necessary condition for knowledge. Hirevlä notes that "the safety condition has been a very popular necessary condition for knowledge of late" (2019: 1). More specifically, safety is introduced to eliminate epistemic luck: "[Safety] seeks to eliminate the kind of luck that is incompatible with knowledge" (ibid.). While there might be more to knowledge than safe belief, safety might be all there is to non-accidental truth. This suggests the following specification of Strong Epistemic Reliability:

[32] The list of safety-proponents includes (but doesn't exhausts): Luper (1984), Sosa (1996), (1999a: 378), (1999b: 142), (2000), Sainsbury (1997), Williamson (2000: ch 7), Pritchard (2002), (2007: 280), (2005), (2009a: 33), (2012: 253), (2015), Goldman & Beddor (2016: 123), and Greco (2012). Many people have argued that safety is better in dealing with skepticism than tracking. See, for example, Sosa (1999b: 146f.), (2000: 15f.), and Pritchard (2002), (2005: 46f.), (2007: 282-285, 290f.), (2009a: 34), (2009b), (2012: 255), (2015: 101).

Epistemic Safety

S's belief that *p* is non-accidentally true if, and only if,
if *S* were to believe that *p*, *p* would be true.

Two comments on Epistemic Safety: First, safety is the contrapositive of the epistemic sensitivity condition. Put in formal terms, safety states $\ulcorner B_S p \mathrel{\Box\!\!\rightarrow} Tp \urcorner$; and sensitivity states $\ulcorner \neg Tp \mathrel{\Box\!\!\rightarrow} \neg B_S p \urcorner$. However, despite being contrapositives, both conditions impose distinct requirements because they are not logically equivalent. Counterfactual conditionals—unlike material conditionals—are in general not logically equivalent to their contrapositives. Formally expressed: $\ulcorner p \mathrel{\Box\!\!\rightarrow} q \urcorner \not\Longleftrightarrow \ulcorner \neg q \mathrel{\Box\!\!\rightarrow} \neg p \urcorner$.[33]

Second, the safety condition is trivially satisfied on the standard semantics of counterfactuals (like adherence; see page 30). This is because the antecedence and the consequence are both true in the actual world. To avoid this problem, "safety theorists give [...] a Nozick-style semantics" for the safety condition (Freitag 2014: 47). Under Nozick-style semantics, the safety condition is satisfied if, and only if, "in (most or) nearly all of the nearby worlds in which *S* believes that *p*, *p* is true" (Becker 2008b: 356).[34]

[33] Sosa notes that epistemic safety and epistemic sensitivity "are not necessarily equivalent, since subjunctive conditionals do not contrapose" (2000: 14). See Williamson (2000: 147), Pritchard (2005: 159), and Ichikawa & Setup (2018).

[34] In fact, most people directly state the safety condition in terms of possible words, rather than subjunctive terms. See, for example, Pritchard (2009a: 34).

Besides being intuitive, Epistemic Safety is often motivated by its ability "to deal with Gettier cases" (Hirvelä 2019: 1).[35] And, in fact, it covers both Clock and Sheep. In Clock, there are many nearby worlds in which *S* looks at the broken clock and believes that it is 3:00 pm when it's actually 3:01 pm. In Sheep, there are many nearby worlds in which *S* looks at the sheep-shaped rock and believes that there is a sheep on the field when, in fact, no sheep is hiding behind the rock. *S*'s beliefs aren't safe and thus only accidentally true.

While many people have suggested a safety-based theory of non-accidental truth, no one has yet proposed a safety-based theory of non-accidental rightness. In analogy to epistemic safety, we can define moral safety as follows: *S*'s action φ is safe if, and only if, if *S* were to perform action φ, φ would be right. Based on this notion of moral safety, one might suggest that an action is reliably right, if and only if, it is safe. In the spirit of Epistemic Safety, we can thus state the following specification of Strong Moral Reliability:

Moral Safety

S's action φ is non-accidentally right if, and only if,
if *S* were to perform action φ, φ would be right.

[35] It is often argued that Epistemic Safety—just as Epistemic Tracking—covers the main Gettier-cases. To give some examples, Pritchard suggests that the "[safety-]principle deals with the paradigm Gettier-style counterexample[s]" (2005: 156; see also 162-173). Pritchard also argues that safety has " been put forward directly as plausible ways of eliminating knowledge-undermining epistemic luck of the kind found in Gettier-style cases" (2015: 99). See also Pritchard (2007: 281), (2012: 254), (2009a: 38f.).

Both comments on Epistemic Safety also apply to Moral Safety. First, moral safety ($\ulcorner A_S\varphi \boxdot\!\!\rightarrow R\varphi\urcorner$) is the contrapositive to moral sensitivity ($\ulcorner \neg R\varphi \boxdot\!\!\rightarrow \neg A_S\varphi\urcorner$). Yet both conditions impose distinct requirements because they aren't logically equivalent. Second, on the standard semantics of counterfactuals, the moral safety condition is trivially satisfied. We can avoid this problem with Nozick-style semantics: S's action φ is safe if, and only if, in most (or all) nearby worlds in which S performs action φ, φ is right.

Moral Safety—like Epistemic Safety—is both intuitive and able to cover (moral) Gettier-cases. In Trolley III, there are many nearby worlds in which S pulls the lever although the rock does not roll on track B and 1000 people die. (There is no Superman in this case.) In Doctor III, there are many nearby worlds in which S prescribes medication M_1 although the patient doesn't drink lemon water and dies. (The water itself is not acidic in this case.) For these reasons, S's actions are not safe and thus only accidentally right.

Having stated and motivated both safety-accounts, we must briefly refine them. Epistemic Safety is usually refined in two ways. First, to cover Nozick's Grandmother case, we must fix the method of belief formation (Becker 2008b, Kelp 2009a, Schafer 2014, Zhao forthcoming). There won't be many nearby worlds in which the grandmother believes that her grandson is all well *on the basis of perception* when, in fact, he is seriously ill. That's why the grandmother's belief is safe and therefore non-accidentally true.

Second, Epistemic Safety on its own can't deal with necessary truths (or truths that are only false in far off worlds).[36] Remember the MATHS case. S forms the belief that $\sqrt[3]{1728} = 12$ by tossing a coin. While the proposition is necessarily true (i.e., there are no worlds in which $\sqrt[3]{1728} \neq 12$), it seems that S's belief is only accidentally true. However, according to Epistemic Safety, S's belief is safe and thus non-accidentally true. There are just no nearby worlds in which S believes that $\sqrt[3]{1728} = 12$ and his belief is false.

To overcome this problem, safety-theorist often introduce a second condition. It is not enough that S's belief via method M is safe, but M itself must also be reliable. To be reliable, M must give the right verdict on many propositions (not just p).[37] This further condition covers MATHS. S forms his belief by tossing a coin. However, tossing a coin is not a reliable method. If one predicts the weather by tossing a coin, one's beliefs will more often than not be false. For this reason, S's belief is only accidentally true. In summary:

Epistemic Safety Refined

S's belief that p via M is non-accidentally true if, and only if,
(a) if S were to believe that p via M, p would be true, and
(b) M is a reliable method.[38]

---

[36] On this point, see (among others) Pritchard (2007: 280, 290), (2012: 256f.), (2015: 101), Roland & Gogburn (2011), Faraci (2019: 1), and Hirevlä (2019: 1).

[37] This solution to deal with necessary truths is proposed by Miscevic (2007), Pritchard (2009a: 34), (2012: 256f.), Hirevlä (2019: 5), as well as Das (n.d.).

[38] Becker argues that we must distinguish the method in condition (a) from the method in condition (b) (2008b: 362). I shall ignore this distinction.

Both refinements also apply to Moral Safety. First, Moral Safety cannot deal with a variation of the Animal case. Consider Animal II: *S* saves a group of strangers. His action deserves praise and is thus non-accidentally right. Yet there are many nearby worlds in which it's wrong to save the strangers because they are terrorists. However, in all those worlds, animals will die if the strangers are not rescued (fill in the details as you like). Being an animal lover, *S* will thus save the strangers even if they are terrorists. *S*'s action is not safe.

We can avoid this problem by fixing *S*'s motives. In the actual world, *S* saves the strangers based on his sympathy for humans. In the counterfactual scenarios, however, *S* action is driven by his love for animals. We can assume that *S*'s sympathy for humans fades when the strangers are terrorists. Keeping S's actual motive fixed therefore explains why his action his safe. When acting on his sympathy for humans, *S* only saves the strangers in those nearby worlds in which they are not terrorists and doing so is right.

Second, Moral Safety can't deal with actions that are necessarily right. In Uncle, *S*'s action is accidentally right but safe. Whenever *S* cares for his orphan nephew because the government pays him, doing so will be right (assuming the action is necessarily right). We can cover Uncle, by requiring that the motive *M* must be reliable. To be reliable, *M* must generally lead to right actions, not just one particular right action φ. It's not reliable to do what the government pays one to do (think about Nazi-governments). In summary:

Moral Safety Refined

S's action φ based on *M* is non-accidentally right if, and only if,
(1) if *S* were to perform φ based on *M*, φ would be right, and
(2) *M* is a reliable motive.

In this section, I have stated, motivated, and refined a specification of Strong Moral Reliability in terms of safety. Both conditions of Moral Safety Refined specify what it is for an action to be reliably right. I have developed Moral Safety Refined as an alternative to Moral Tracking Refined—just as safety-based accounts of knowledge are developed as alternatives to Nozick's tracking-theory of knowledge. In the next section, however, I argue that Moral Safety Refined fails for the same reasons as Moral Tracking Refined.

2.5 Against Safety

To show how Moral Safety Refined fails, I first demonstrate the inadequacy of Epistemic Safety Refined. Let's start with the original Barns case. *S* walks through fake-barn-county. Looking at the only real barn, *S* forms the belief that there is a barn on the field in front of him. *S*'s belief is only-accidentally true. At this stage, Epistemic Safety Refined can easily explain this verdict because *S*'s belief is not safe. There are many nearby worlds in which *S* forms the same belief while standing in front of a fake barn.

Again, we must distinguish two scenarios in which *S*'s belief is not safe. *S* could have looked at a fake barn on a different field,[39] or a fake barn could have been on the same field as the real barn before. Barns II blocks both scenarios (see pages 39f.). We don't assess *S*'s belief that there is a barn on the field in front of him, but his belief that there is a barn on field f_3. Moreover, we assume that no fake barn can be erected on field f_3 because the ground is too rocky. Still, *S*'s belief is only accidentally true.

In this modified case, *S*'s belief turns out to be safe. In most nearby worlds in which *S* looks at field f_3 and continues to believe that there is a barn on this field, his belief will be true. Furthermore, perception is (generally speaking) a reliable method to form beliefs. Both conditions of Epistemic Safety Refined are therefore satisfied, but *S*'s belief is only accidentally true. This suggests that safety is not sufficient for non-accidental truth. Note that Barns II is not an isolated counterexample to Epistemic Safety Refined.

For instance, *S*'s beliefs in Clock II and Sheep II are accidentally true but safe. In Clock II, *S* believes that it is 3:00 pm only in those worlds in which it is actually 3:00 pm because the benevolent demon would have otherwise changed the clock face. In Sheep II, *S* believes

[39] For example, Beddor & Pavese argue that "[*S*] could have easily looked out the window a moment earlier; had he done so, he would have formed the same belief, but it would have been false" (forthcoming). Based on considerations like these, Gaultier concludes: "[*S*] does not know that there is a barn in front of him because his true belief that there is a barn in front of him is *unsafe*" (2014: 478). For a similar line of argument, see also Pritchard (2005: 161), (2009a: 38f.), (2012: 254), (2015: 102f.) as well as Kelp (2009a: 28).

that there is a sheep on the field only in those worlds in which there is actually a sheep on the field. For hadn't there been a sheep, the sheep-shaped rock would have been removed. In both cases, *S* bases his beliefs on perception, which is generally a reliable method.

Modified Gettier-cases like Barns II, Clock II, and Sheep II are not only cited against Nozick's tracking-theory of knowledge. They are also used as counterexamples to theories of knowledge as safe belief, showing that safety is only a necessary but not sufficient condition for knowledge: "[T]here is [...] more to knowledge than safe belief" (Pritchard 2009a: 34).[40] I have argued that modified Gettier-case are also counterexamples to Epistemic Safety Refined. They show that safety is not sufficient for non-accidental truth.

Structurally identical considerations speak against Moral Safety Refined. The moral analogue to Barns is the Beggars case. *S* walks through fake-beggar-county. Standing in front of the only real beggar, *S* donates money to the person on the slot in front of him. His action is only accidentally right. At this stage, Moral Safety Refined can still explain the verdict. *S*'s action is not safe because there are many nearby worlds in which *S* donates money to the person on the slot in front of him when that person is a fake beggar.

[40] Pritchard (2009a) is not the only Person to reject the sufficiency of safety for knowledge. See also Hiller & Neta (2007: 307), Pritchard (2009a: 41), (2012), (2015: 103), Freitag (2014: 45), Schafer (2014: 384), Hales (2016: 490), Engel (2019), Hirevlä (2019: 2), Beddor & Pavese (forthcoming), and Das (n.d.). For further objections against Epistemic Safety, see DeRose (2002), Greco (2003), Kvanvig (2004), Murphy (2005: 333f.), as well as Kelp (2009a: 25f.), (2009b).

Again, we must distinguish two scenarios in which *S*'s action is not safe. *S* could have stood in front of a fake beggar on a different slot, or a fake beggar could have been on the slot of the real beggar. Both scenarios can be blocked with Beggars II: We assess *S*'s action under the description donating-money-to-the-person-on-slot-s_3. Moreover, if no real beggar sits on slot s_3, the law demands that a police officer blocks the slot to ensure that no fake beggar is sitting down. Still, S's donation is only accidentally right.

In this modified beggar case, *S*'s action turns out to be safe. In most nearby worlds in which *S* donates money to the person on slot s_3, the action will be right because *S* only donates to the real beggar, not the police officer. We can further assume that *S*'s donation is based on compassion, a generally reliable motive. *S*'s donation thus satisfies both conditions of Moral Safety Refined despite being only accidentally right. Safety is not sufficient for non-accidental rightness. Beggars II is also not an isolated counterexample.

For instance, consider Trolley IV: *S* pulls the lever, thereby diverting the trolley to track *B*. A rock stops the trolley, saving the 1000 people on track *B*. Had there been no rock, Superman would have stopped the trolley instead. *S*'s action is only accidentally right but safe. *S* pulls the lever in those worlds in which he sees less than five people on track *B*. In all these worlds, his action is right because the 1000 people are either protected by the rock or Superman. Besides, *S*'s rational considerations are a reliable motive.

Consider Doctor IV: S prescribes medication M_1. After taking the medication, the patient starts to sweat heavily and drinks lemon water. The acid from the lemon chemically reacts with M_1 thereby curing the patient. But since we assume that the water is acidic too, really any drink would have worked. S's prescription is accidentally right but safe. In most nearby worlds in which S writes the prescription, the patient will be cured because he reliably consumes the acid. Moreover, S's care for his patients is a reliable motive.

In sum, modified moral Gettier-cases like Beggars II, Trolley IV, Doctor IV not only refute the tracking-theory of non-accidental rightness, they also work as counterexamples to the safety-account. S's actions in Beggars II, Trolley IV, Doctor IV satisfy both conditions of Moral Safety Refined while only being accidentally right. I contend that there are many more such modified moral Gettier-cases. We can therefore confidently conclude that safety is not sufficient for non-accidental rightness.

On a larger picture, I have developed Moral Tracking Refined and Moral Safety Refined as two plausible specifications of Strong Moral Reliability. As we have seen, both specifications are subject to the same set of counterexamples: Beggars II, Trolley IV, Doctor IV. What are the implications for Strong Moral Reliability in general? Should we search for better specifications? Should we reject the modal account altogether? Or should we only restrict it? In the next section, I advocate for a restriction of the modal account.

2.6 Restriction

In this chapter, I have first stated and motivated the modal account of non-accidental rightness (Strong Moral Reliability). The account says that an action is non-accidentally right *if*, and only if, it is reliably right (§ 2.1). I have then specified the notion of reliable rightness in terms of tracking and safety (§§ 2.2 and 2.4). Finally, I have argued that both specifications are not sufficient for non-accidental rightness (§§ 2.3 and 2.5). In this section, I discuss three possible implications for the modal account.

First, one might argue that we should search for a better specification of Strong Moral Reliability. Maybe there is a specification of reliable rightness that avoids the problems of tracking and safety.[41] More precisely, there might be a specification of reliable rightness that is both necessary *and* sufficient for non-accidental rightness. On this view, the counterexamples only show that Moral Tracking Refined and Moral Safety Refined are not the right specifications of Strong Moral Reliability. They don't infest the modal account.

[41] In fact, at this point, one might again draw on resources in epistemology. Riggs has recently developed the idea of a mere coincidence: "To say that the coincidence of A and B is 'mere' is to say that A and B are completely independent of one another—that one had 'nothing to do with' the other" (2014: 635f.). Riggs then uses the idea of a mere coincidence to clarify our notion of epistemic luck: "[I]t is a mere coincidence that (a) S has a [...] belief that p and (b) that p is true" (2014: 637). In the same spirit, one might specify the notion of reliable rightness: It is not a mere coincidence that (a) S performs action φ and (b) that φ is right. Whether this alternative specification really avoids the problems of tracking and safety depends on the details.

While it is certainly possible that there is a better specification of Strong Moral Reliability, I do not think that it is very likely. I have developed Moral Tracking Refined and Moral Safety Refined in analogy to their epistemic counterparts: Epistemic Tracking Refined and Epistemic Safety Refined. Both epistemic accounts are the result of decades-long research on epistemic luck. If tracking and safety are the best specifications of Strong Epistemic Reliability, we can assume the same for Strong Moral Reliability.

Second, on the other extreme, one might argue that the counterexamples against both specifications of the modal account show that it is generally confused to think about non-accidental rightness in modal terms. We should reject Strong Moral Reliability altogether. As I have noted before (see page 35), some people actually pursue this route (though for different reasons): "[H]ow the agent might have acted counterfactually is not relevant to [...] moral worth," and thus non-accidental rightness (Markovits 2010: 215).

I think this second response to the counterexamples is premature. It throws the baby out with the bath water. The counterexamples only show that tracking and safety are not sufficient for non-accidental rightness; they don't show that tracking and safety are not necessary. Accordingly, the counterexamples don't give us any reason to reject the modal account altogether. Reliable rightness might still be necessary. And, in fact, it seems intuitive that the relevant notion of non-accidental rightness has some modal dimension.

Third, and finally, I suggest to restrict Strong Moral Reliability. On this view, reliable rightness is necessary but not sufficient for non-accidental rightness. I think this is the right conclusion to be drawn from the counterexamples against tracking and safety. On the one hand, we don't have to speculate about better specifications of Strong Moral Reliability. On the other hand, we do not prematurely reject the modal account altogether. I therefore propose the following modal condition on non-accidental rightness:

Moral Reliability

S's action φ is non-accidentally right *only if*
S's action φ is reliably right

It should be noted that the same considerations apply to the epistemic domain. The counterexamples against Epistemic Tracking Refined and Epistemic Safety Refined show that tracking and safety are not sufficient for non-accidental truth. At this point, one might search for better specifications of Strong Epistemic Reliability, or reject the modal account altogether. I think that both options are unattractive. We should instead restrict Strong Epistemic Reliability. This gives us the modal condition on non-accidental truth:

Epistemic Reliability

S's belief that p is non-accidentally true *only if*
S's belief that p is reliably true.

At the beginning of this chapter (§ 2.1), I have motivated Strong Moral Reliability and Strong Epistemic Reliability with two considerations: they are intuitive and cover many cases. Both considerations transfer to their restricted counterparts. First, Lackey (2006, 2008) has argued that the modal account of non-accidental events in general should be restricted. On her theory, counterfactual robustness is not sufficient for non-accidentality: $\langle e \rangle$ occurs non-accidentally *only if* the occurrence of $\langle e \rangle$ is reliable.

Lackey motivates this restriction with the following Treasure case: S plants roses on the only patch of soil on an otherwise rocky island. While digging, S finds a treasure. Lackey notes that the "discovery of the buried treasure is a lucky event" (2006: 286). However, the treasure could not have been buried somewhere else on the island because all other patches are rocky. Lackey concludes that S finding the treasure "also occurs in a wide class of the nearest possible worlds" (ibid.). The event is accidental but reliable.

Moral Reliability and Epistemic Reliability are intuitive because they are instances of this restricted modal condition of non-accidental events in general. To be more specific, Moral Reliability states that events of the form $\langle S$'s action φ being right$\rangle$ occur non-accidentally only if their occurrence of $\langle S$'s action φ being right$\rangle$ is reliable. In the same way, Epistemic Reliability claims that events of the form $\langle S$'s belief that p being true$\rangle$ occur non-accidentally only if their occurrence of $\langle S$'s belief that p being true$\rangle$ is reliable.

Second, Moral Reliability and Epistemic Reliability cover the same cases as their unrestricted counterparts (at least the cases discussed in § 2.1). I've introduced both modal accounts to explain the *lack* of moral worth and non-accidental rightness, on the one hand, and the *lack* of knowledge and non-accidental truth, on the other hand. However, to explain this lack, we need reliable rightness and reliable truth only as necessary conditions. The violation of these conditions explains the accidental rightness and truth.

In closing this chapter, I should note that we could again specify Moral Reliability in terms of tracking or safety. However, I do not favour any particular specification (as long as it is restricted). The sole purpose of developing the notions of tracking and safety was to show that the best specifications of the unrestricted modal account are not sufficient for non-accidental rightness. I therefore leave Moral Reliability unspecified. (The reader might choose his or her favourite specification from the epistemic domain.)

3 The Control Condition

In this chapter, I first state and motivate the control condition for non-accidental rightness (§ 3.1). I then introduce a variation of the problem of moral luck as an objection to the control condition. The objection argues that we usually lack the relevant type of control (§ 3.2). Next, I refute this objection by weakening the underlying notion of control (§ 3.3). Finally, I restrict the control condition, arguing that it is not sufficient for non-accidental rightness; neither individually nor in combination with the modal condition (§ 3.4).

3.1 Motivation

To state and motivate another necessary condition on non-accidental rightness, I build on the results of the last chapter. There I have argued that reliable rightness is not sufficient for non-accidental rightness. *S*'s actions in Beggars II, Trolley IV, and Doctor IV are reliably right, but their rightness is only accidental. What goes wrong in these cases? To answer this question, it is again helpful to first examine the epistemic literature on the analogue question: Why is reliable truth not sufficient for non-accidental truth?

Virtue epistemologists have argued that knowledge is a cognitive achievement for which one deserves credit. Here are some examples: Greco suggests that attributing knowledge to someone "is to say that his believing the truth can be credited to him" (2003: 112). Riggs claims that "knowledge is an achievement, something for which the agent can take credit" (2007: 277). Carter et al. propose that "knowledge is type-identical with a particular species of cognitive achievement" (forthcoming)—to cite just a few.[42]

Knowledge excludes luck. If knowledge is some sort of cognitive achievement, we can infer that this achievement excludes luck, too. Indeed, we wouldn't want to credit someone for something which is only a matter of luck. More specifically, we don't think someone deserves credit for a true belief if that belief is true only by accident. In fact, Pritchard turns the argument on its head, arguing that knowledge excludes luck because it is a cognitive achievement and achievements exclude luck (2005: 2).[43]

Thinking about knowledge as a cognitive achievement helps us to isolate a specific type of epistemic luck. One can only take credit for the truth of a belief if the belief's truth has "at least *something* to do with the agent" (Pritchard 2005: 188).[44] Conversely, it seems that a

[42] On this issue, see also Pritchard (2005: 1f.), Greco (2007), (2010), (2012), Riggs (2007: 331), (2009b), (2014: 627), Sosa (2007), (2009), and Gaultier (2014: 477).

[43] In this context, Riggs states: "[T]here are deep conceptual connections between the notion of credit [...] and the notion of luck" (2007: 336; see also 341).

[44] The importance of the agent for the presence or absence of luck is also discussed by Pritchard (2005: 185), (2015: 103), Riggs (2007: 332-335), Gaultier (2014: 477) Schafer (2014: 381f.), as well as (Das n.d.).

belief, whose truth has (in a relevant sense) nothing to do with the agent (but with the environment), is only accidentally true. Using Breyers terminology, we can say that a "belief is agentially lucky" if its truth cannot be credited to the agent (2010: 143).

With this pre-theoretic notion of agential luck, we can explain why reliable truth isn't sufficient for non-accidental truth. In modified Gettier-cases like Barns II, Clock II, and Sheep II, *S*'s beliefs are reliably true, and yet their truth appears to be accidental. This fact can be explained with the presence of agential luck. In the modified cases, the truth of *S*'s beliefs is ensured by the environment, not *S*. *S* therefore cannot take credit for the truth of his beliefs. Epistemic Reliability doesn't exclude agential luck.

The notion of agential luck can be analysed in different ways. Some analyses are better than others. For instance, one might suggest that an agent can be credited for the truth of his belief just in case he exhibits the relevant cognitive *abilities* when forming the true belief. This analysis fails. For it seems that in Barns II—and maybe also in Clock II, and Sheep II—*S* exhibits the relevant cognitive abilities. When *S* looks at the barn (the clock, the rock), he forms his true beliefs on the basis of perceptual vision.[45]

[45] In fact, many people have argued that *S* exhibits his cognitive abilities in barns-cases. To give an example, Pritchard states that "we should treat [*S*] as exhibiting his ordinary cognitive ability to spot barns" (2012: 269). The reason for this: "[I]n the barn facade case the [*S*] is genuinely looking at a real barn" (2015: 105). See also Pritchard (2009a) and Gaultier (2014: 477).

Rather than analysing agential luck in terms of cognitive abilities, I propose to cover agential luck in terms of control.[46] Riggs states: "[K]nowing that p depends upon one's *having control* [...] over the fact that one came to hold a true belief" (2007: 341; my emphasis).[47] Even if the control over the truth of one's belief is not sufficient for knowledge, it seems that the lack thereof accounts for agential luck. An agent cannot take credit of a true belief if he has no control over his belief being true. We can thus state the following condition:

Epistemic Control

S's belief that p is non-accidentally true only if S has control over the truth of his belief that p.[48]

At this point, I should highlight two potential misunderstandings. First, Epistemic Control only requires that we *have* control over the truth of our beliefs. However, it does not require that we also *exercise* that control. Generally speaking, one can have control over something without exercising that control. To give an example, the parliament has control over the government. Yet it might choose not to exercise that control by letting the government do whatever it wants. (I return to this distinction in § 4, pages 100f.)

46 Here I follow the suggestion of Broncano-Berrocal: "The agent-relative sense of risk [...] can be [...] defined as lack of control over the event" (2015: 11).

47 In a later paper, Riggs explicitly analyses knowledge in terms of control: "S knows that p if and only if S's believing the truth about p is an outcome that is/was sufficiently under S's control" (2014: 628). The relation between knowledge and control is also discussed by Axtell (2003) and Coffman (2007).

48 For example, Engel writes that "epistemic luck [can] be cashed out in terms of a lack of control condition" (2019: § 1).

Second, to satisfy the control condition, we must have control over the *truth* of our belief that p. Epistemic Control does not require that we have control over our *belief* that p. Epistemic involuntarists argue that we have no (voluntary) control over any specific belief. Even if true, this does not imply that all of our beliefs are merely accidentally true. It is conceptually possible that one controls the truth of a belief without controlling the belief itself. (I say more about the control over truth in §§ 3.2 and 3.3.)

Control is highly sensitive to the environment (much more so than abilities). Placed in a bad environment, an agent loses the control over the truth of his beliefs. I suggest that modified Gettier-cases like Barns II, Clock II, and Sheep II describe bad environments. The surroundings of fake-barn-county, the presence of a broken clock, or the existence of a sheep-shaped rock might not impair S's cognitive abilities, but they plausibly undermine S's control over the truth of his beliefs (for details, see page 91).

To summarise, Epistemic Control excludes agential luck (unlike Epistemic Reliability). The truth of a belief is agentially lucky if it cannot be credited to the agent. We can only credit the truth of a belief to an agent if the agent controls that truth. Epistemic Control can thus explain why modified Gettier-cases like Barns II, Clock II, and Sheep II are only accidentally true (despite being reliably true). They are only accidentally true because the truth is agentially lucky; the agent doesn't control the truth of his beliefs.

To be sure, Epistemic Control is an unorthodox view in the epistemic literature. However, it provides a helpful template to think about non-accidental rightness. Many people argue that knowledge is some sort of cognitive achievement. By analogy, I suggest that moral worth is some sort of moral achievement for which the agent deserves to be credited. If a specific action is a moral achievement, it deserves moral praise. And conversely, if a specific action deserves moral praise, it plausibly is a moral achievement.

Moral achievements exclude luck. More specifically, moral achievements, like all other achievements, exclude agential luck. We have seen that a belief is agentially lucky if its truth cannot be credited to the agent. Likewise, we might say that an action is agentially lucky if its rightness cannot be credited to the agent (but is merely due to environmental features). We clearly don't think that the action being right is a moral achievement on the side of the agent when the action's rightness can't be credited to him.

This pre-theoretic notion of agential luck can explain why reliable rightness isn't sufficient for non-accidental rightness. In modified moral Gettier-cases like Beggars II, Trolley IV, and Doctor IV, *S*'s actions are accidentally right, despite their rightness being reliable. We can explain this fact with the presence of agential luck. In the modified cases, the rightness of *S*'s actions is mainly ensured by the environment. *S* therefore cannot take credit for the rightness. Moral Reliability doesn't exclude agential luck.

Learning from the epistemic domain, I think we shouldn't analyse this type of agential luck in terms of moral abilities. For one might plausibly argue that *S* exhibits the relevant moral abilities when performing his actions in Beggars II, Trolley IV, and Doctor IV. In these modified cases, *S* is compassionate, rational, and he cares for his patients. In analogy to Epistemic Control, I suggest that an action is agentially lucky if the agent doesn't control its rightness. I therefore propose the following necessary condition:

Moral Control

S's action φ is non-accidentally right only if *S* has control over the rightness of his action φ.

Like Epistemic Control, Moral Control must be clarified in two ways. First, Moral Control only requires that we *have* control over the action's rightness. It does not require that we *exercise* that control. Second, to satisfy Moral Control, it must be in our control that the action is *right*. It's not enough that we control the *action* itself. This is important because *S*'s actions in Beggars II, Trolley IV, and Doctor IV are only accidentally right. But we can plausibly assume that *S* has control over these actions.

S doesn't have control over the rightness of his actions in Beggars II, Trolley IV, and Doctor IV. Control is highly sensitive to the environment. I suggest that the modified Gettier-cases describe bad environments. The constitution of fake-beggar-county, the events

on track B, and the presence of the patient's rare medical condition plausibly undermine S's control over the rightness (though they don't impair his moral abilities). That's why S's actions are agentially lucky (for more details, see pages 88-90).

In summary, Moral Control excludes agential luck (unlike Moral Reliability). An action is agentially lucky if its rightness cannot be credited to the agent. We can only credit the rightness of an action to the agent if the agent controls that rightness. Moral Control can therefore explain why S's actions in the modified moral Gettier-cases like Beggars II, Trolley IV, and Doctor IV are accidentally right (despite being reliable). The actions are only accidentally right because S doesn't control their rightness.

Not only do both control conditions (moral and epistemic) account for the modified Gettier-cases, they are also intuitive. In addition to the modal condition of non-accidental events in general (page 60), people have also defended a control condition. Riggs, for example, suggests that "E is lucky for S [if] E is (too far) out of S's control" (2009a: 218).[49] Based on this suggestion, we can state the following condition on non-accidental events: An event $\langle e \rangle$ occurs non-accidentally only if S has control over the occurrence of $\langle e \rangle$.

[49] See also Riggs (2007: 215-220). To be fair, such a control condition on non-accidental events has been criticised by several authors. See Latus (2000), Hurley (2003: 111), Pritchard (2005: 127), (2006), (2014: 611-614), Lackey (2008: 256), and Hales (2016). More commonly, people have argued that luck excludes control (i.e., control is sufficient for non-accidentality): "An event is lucky for you only if the event is beyond your control" (Coffman 2009: 499).

Moral Control and Epistemic Control are intuitive because they are instances of this control condition of accidental events in general. We can interpret Moral Control as stating: $\langle S$'s action φ is right$\rangle$ occurs non-accidentally only if S has control over the occurrence of $\langle S$'s action φ is right$\rangle$. Likewise, we can interpret Epistemic Control as stating: $\langle S$'s belief that p is true$\rangle$ occurs non-accidentally only if S has control over the occurrence of $\langle S$'s belief that p is true$\rangle$. I thus claim that both control conditions are well motivated.

In this section, I have stated and motivated the control condition on non-accidental rightness (and truth). I have suggested that S's actions in the modified moral Gettier-cases are only accidentally right because S doesn't control their rightness. I readily admit that the phrase control-over-the-action's-rightness is vague. In the next two sections, I therefore clarify what it takes to have control over the rightness of one's actions (and the truth of one's beliefs).[50] I begin by discussing a variation of the problem of moral luck.

[50] Note that I do not attempt to clarify the notion of control in general. That is to say, I do not attempt to clarify what it is for a subject S to have control over some event $\langle e\rangle$. We can clarify the more specific notion of control-over-the-action's-rightness without clarifying the general notion of control. However, for a brief clarification see Riggs: "S controls E iff S brought E about (where this implies both that E was the result of the application of S's powers, abilities, or skills, and E was not inadvertent with respect to S)" (2009a: 218). Note that a full discussion of control in general would also have to make reference to the Principle of Alternative Possibilities and Frankfurt-cases. Such a full discussion lies outside the scope of this thesis.

3.2 Skepticism and Moral Luck

In this section, I first state the problem of epistemic luck. Drawing on this skeptical problem, I formulate a variation of the problem of moral luck. This variation can be interpreted as an objection to the control condition. First off, skeptical problems question the availability of knowledge by arguing that one or more necessary conditions for knowledge are rarely (or never) satisfied. The problem of epistemic luck, claims that it is difficult (or impossible) to avoid epistemic luck (Latus 2000: 153f.). Put schematically:

The Problem of Epistemic Luck

(P1) If there is knowledge, there is no epistemic luck.
(P2) There is epistemic luck.
(C) *Therefore*, there is little (or no) knowledge.

The first premise just is a rough version of Epistemic Incompatibility. *S*'s belief that *p* amounts to knowledge only if *S*'s belief that *p* is non-accidentally true. As I have argued in the introduction (§ 1), this principle is false for benign luck. Benign luck includes status luck, capacity luck, and evidence luck (see pages 10f.). However, the first premise is necessarily true—and universally accepted (see footnote 14)—for malicious luck. For the argument to be sound, we must fix both premises to malicious luck.

There are two potential arguments for the second premise. First, one could argue that most (or all) of our beliefs are only accidentally true because they are not reliably true. However, this argument isn't plausible. It seems that many of our beliefs satisfy Epistemic Reliability. In fact, many authors have highlighted the anti-skeptical advantages of the safety principle (see footnote 32). Based on these considerations, Engel concludes that "knowledge destroying luck [...] is far from ubiquitous" (2019: § 2).

Engel's conclusion is premature. For one could secondly argue that most (or all) of our beliefs are only accidentally true because we generally lack control over their truth. And, in fact, Nagel thinks that "epistemological skepticism arises from consideration of the respects in which our beliefs and their relation to reality depend on factors beyond our control" (1979: 27). In other words, Nagel suggests that most (if not all) beliefs violate Epistemic Control—we usually do not control the truth of our beliefs.

Unfortunately, Nagel doesn't specify why he thinks that we don't have control over the truth of our beliefs. For the time being, let us assume that Nagel's suggestion is at least initially plausible. As we will later see, Nagel specifies why he thinks that we don't control the rightness of our actions. I use these considerations to retrospectively motivate his claim that we don't control the truth of our beliefs (see page 81). For now, let us assume that both premises of the problem of epistemic luck are true.

Insofar as the problem of epistemic luck is valid, we have three options. First, we could bite the bullet and accept the skeptical conclusion: We have much less knowledge than we think. I don't think this option is attractive. Second, we could reject the first premise. However, Epistemic Incompatibility is nearly universally accepted. Third, and most plausibly, we could reject the second premise by dismissing Epistemic Control. The problem of epistemic luck turns out to be an objection against Epistemic Control.

We can raise an analogous objection against Moral Control (this time with more details). Nagel (1979) and Williams (1981) famously introduce the problem of moral luck—a problem which they, and others, take to be analogous to the problem of epistemic luck.[51] We can summarise this analogy as follows: Knowledge excludes epistemic luck; morality excludes moral luck. Epistemic luck is pervasive; moral luck is pervasive. Thus there is little knowledge; thus there is little morality. The problem in its schematic form:

The Problem of Moral Luck

(P1) If there is morality, there is no moral luck.
(P2) There is moral luck.
(C) *Therefore*, there is little (or no) morality.

[51] Latus writes that "just as we are inclined to believe that luck cannot play a role in determining a person's moral standing, so epistemologists have been inclined to think that luck cannot play a role in instances of knowledge" (2000: 161). On this, see also Nagel (1979: 27), Williams (1993: 242), Axtell (2003: 333), Pritchard (2005: 10), (2006: 2, 21-23), as well as Ho (forthcoming).

The term 'morality' in the first premise operates as a placeholder for a wide variety of more specific moral judgments. Traditionally, the problem of moral luck has been specified to judgments about responsibility, rational justification, blame, as well as a person's moral character.[52] However, the problem of moral luck can also be specified to judgments about the positive moral worth of an action. In this context, I read the term 'moral luck' as 'accidental rightness'. This gives us the following variation of the problem of moral luck:

The Problem of Moral Worth

(P1) If an action φ has moral worth, the action φ is non-accidentally right.

(P2) Most (or all) actions φ are accidentally right.

(C) *Therefore*, most (or all) actions φ don't have moral worth.

The first premise just is Moral Incompatibility. As I have suggested in the introduction (§ 1), Moral Incompatibility is false for all types of benign luck—this includes status luck, capacity luck, and evidence luck (see pages 6f.).[53] However, Moral Incompatibility is necessarily true for malicious luck. On this reading, the principle is universally accepted (see footnote 12). The problem of moral worth (like the problem of epistemic luck) is only sound if we fix both premises to luck in the malicious sense.

[52] Nelkin writes: "[I]t is important to recognize that there are many different kinds of moral assessment" (2019: § 3). See also Nagel (1979: 25), Williams (1981: 22), Richards (1986), Greco (1995: 91), and Latus (2000: 150), (2019).

[53] Nagel famously distinguishes between four types of moral luck (1979: 28).

Again, there are two potential arguments for the second premise. First, we could argue that most (or all) of our right actions are only accidentally right because they are not reliably right. In other words, most (or all) right actions violate Moral Reliability. However, similar to its epistemic equivalent, this argument isn't very convincing. There is simply no good reason to think that our actions can't be reliably right. On the contrary, we might think that many virtuous agents often perform actions that are reliably right.

Second, one might argue that most (or all) of our actions are only accidentally right because we lack control over their rightness. Nagel often suggests that the general problem of moral luck originates in a lack of control: "If the condition of control is consistently applied, it threatens to erode most [...] moral assessments [...]. The things for which people are morally judged are determined [...] by what is beyond their control" (1979: 26). These considerations also apply to judgments about the moral worth of actions.

Latus gives the following summary: "(i) Constitutive Luck: the luck involved in a person's having the inclinations, capacities and temperament that she does. (ii) Circumstantial Luck: the luck involved in 'the kind of problems and situations one faces.' (iii) Causal Luck: 'luck in how one is determined by antecedent circumstances.' (iv) Resultant Luck: 'luck in the way one's actions and projects turn out"' (2000: 152). For a further discussion, see Nelkin (2019), and Ho (forthcoming). It is important to note that not all four types of luck are malicious. Constitutive luck and causal luck are close to what I call capacity luck (which is benign). Resultant luck is only malicious if we assume a broadly consequentialist theory of the right. For if the action's rightness does not depend on its results, it won't be accidentally right when its results are accidental. Circumstantial luck is only malicious when the circumstances are not counterfactually robust. I don't think Nagel's distinction between the four types of luck is joint-carving, and so I ignore it.

But why should we think, that we rarely (or never) control the rightness of our actions? Nagel uses the famous example of two reckless drivers to give a case-based argument why we rarely (or never) control the *wrongness* of our actions (1979: 25). I won't discuss Nagel's example here,[54] but I think we can create an analogous case-based argument for the claim that we rarely (or never) control the *rightness* of our actions. While Nagel contrasts two reckless drivers, I contrast two skilled hunters. Here are my cases:

Hunter I: S is a skilled hunter, responsible for managing the number of wolfs in an area A. If the wolf population in A gets too large, the wolfs can't find enough natural prey (mainly deer) and start attacking the people in the region. However, if there are not enough wolfs in A, the deer population will explode, thereby destroying the eco-system in A. Being a professional hunter, S tries to maintain a balance between wolfs and deer. One day, S shoots a wolf. His action was right because there were too many wolfs in A.

Hunter II: As before, S is a skilled and professional hunter. It is right to shoot a wolf only if the wolf population in A gets too large in comparison to the deer population. One day, S shoots a wolf based on the information that there are too many wolfs in area A.

54 In short, S drives recklessly and hits a pedestrian (unlike the second driver). Since S doesn't control all the wrong-making features of his action (especially the presence of the pedestrian), S lacks control over the action's wrongness, and so his action is only accidentally wrong. This argument draws on the principle that one controls the wrongness of an action if, and only if, one controls the action's wrong-making features. For a further discussion, see Latus (2000: 158f.), Pritchard (2005: 257-259), (2006: 4), and Riggs (2007: 332).

However, just as *S* shoots the wolf, a small asteroid kills a large part of the remaining wolf population. As a consequence, the deer population explodes and destroys the ecosystem. In this case, it seems that *S* (objectively) shouldn't have shot the wolf.

The contrast between HUNTER I and HUNTER II achieves two things. First, it isolates one particular feature which makes the difference between right and wrong, namely the presence (or absence) of the asteroid. Second, the contrast makes it plausible that *S* doesn't control this feature in either case. With a view to HUNTER I, we can conclude that *S* doesn't control (all) right-making features of his actions. From this one might infer that *S* doesn't control the action's rightness. This inference draws on the following principle:

Strong Control

S has control over the rightness of an action φ if, and only if,
S controls the right-making features of φ.

There is no doubt, Strong Control is highly intuitive. However, given this principle, it seems that we rarely (or never) control the rightness of our action because we rarely control (all) the right-making features of our actions. This claim is particularly plausible when the action's rightness depends on features in the agent's environment. As Williams notes: "[W]hat is not in the domain of the self is not in its control, and so is subject to luck" (1981: 20). These considerations imply that most (or all) right actions are only accidentally right.

(On a side note, we could employ the analogous line of reasoning to motivate the claim that we rarely (or never) control the truth of our beliefs. One might argue that we control the truth of a belief if, and only if, we control its true-making features. However, in most cases, we don't have control over the true-making features of our beliefs. To give an example, I believe that the earth rotates around the sun without controlling this fact. The same reasoning applies to mathematical propositions like $\sqrt[3]{1728} = 12$.)

I have motivated both premises of the problem of moral worth. Assuming that the problem is valid, we have three options. First, we could bite the bullet and accept the conclusion that (almost) no action deserves moral praise. However, this is just as unattractive as accepting the skeptical conclusion in the epistemic domain. Second, we could reject the first premise. However, this amounts to rejecting Moral Incompatibility, which is universally accepted. Third, and finally, we could reject the second premise.

I take it that the third option is most plausible. However, in order to reject the second premise, we must dismiss one of the supporting assumptions. My argument for the second premise rests itself on three premises. First, an action is non-accidentally right only if the agent controls its rightness (Moral Control). Second, we control an action's rightness if, and only if, we control its right-making features (Strong Control). Third, we rarely (or never) control the right-making features of our actions. More explicitly:

The Supporting Argument

(P1) An action φ is non-accidentally right only if the agent controls the rightness of φ.

(P2) An agent controls the rightness of an action φ if, and only if, he controls the right-making features of φ.

(P3) For most (or all) actions φ, the agent does not control the right-making features of φ.

(C) *Therefore*, most (or all) actions φ are accidentally right.

Which premise of the supporting argument should we dismiss? One might argue that the plausibility of the third premise highly depends on one's first-order moral theory. For instance, Kantians might argue that an action is right if, and only if, its maxim is universalisable. We have full control over our maxims (though see page 26). Even if this is true, the third premise seems plausible for all moral theories which think that the agent's environment (e.g., his consequences) determines whether his action is morally right.

To solve the problem of moral worth, we are therefore left with two options. Either we dismiss the first or the second premise of the supporting argument; either we dismiss Moral Control or Strong Control. If we maintain Strong Control, we must "conclude [...] that the condition of control is false" (Nagel 1979: 26). This is the objection against Moral Control. In the next section, I defend Moral Control, by weakening (i.e., dismissing) Strong Control. In this way, we can keep Moral Control and solve the problem of moral worth.

3.3 Weak Control

In this section, I defend Moral Control by weakening Strong Control. To begin with, I discuss three straightforward but ultimately unsuccessful ways to weaken Strong Control. First, in my discussion of Strong Control, I have implicitly assumed that it requires *direct* control over all right-making features. If we don't have that control, we also don't have control over the rightness of our actions. Against this background, one might suggest that it's sufficient to *indirectly* control the right-making features of our actions:

Indirect Control

S has control over the rightness of action φ if, and only if, *S* directly or indirectly controls the right-making features of φ.

Indirect Control is an intuitive weakening of Strong Control. At this point, one would have to say more about the distinction between direct and indirect control. Direct control might be analysed in terms of immediate causal influence, whereas indirect control only requires mediate causal influence. While we lack direct control over the right-making features in our environment, it seems plausible that we often have indirect control. If true, this solves the problem of moral worth without rejecting Moral Control.

While it seems plausible that we indirectly control *some* right-making features in the environment, I don't think that we have indirect control over *all* right-making features. In HUNTER I and HUNTER II, the difference between right and wrong was due to an asteroid. Clearly, *S* has no indirect control (in any relevant sense) over the presence or absence of the asteroid. Plausibly, there are always some right-making features over which the agent doesn't even have indirect control. Indirect Control is not weak enough.

Second, one might suggest that we only need (direct) control over *some* rather than *all* right-making features—see Partial Control below. Without specifying the number of right-making features required, it seems that Partial Control avoids the difficulties of Indirect Control. While *S* has no control over the asteroid, *S* controls at least some right-making features, e.g., the number of deer. In fact, there are usually some right-making features over which we do have control. Partial Control thus solves the problem of moral worth.

Partial Control

S has control over the rightness of action φ if, and only if, *S* controls some right-making features of φ.

However, Partial Control is satisfied by *S*'s actions in the modified Gettier-cases. In BARNS II, *S* controls the happiness of the beggar by giving him the money. In TROLLEY IV, *S* controls the survival of the five people on track *A* by pulling the lever. In DOCTOR IV, *S*

controls that the patient consumes medication M_1 by writing the prescription. The happiness of the beggar, the survival of the five people on track A, and the patient's consumption of medication M_1 contribute to the (objective) rightness of S's actions.

If S's actions in the modified Gettier-cases satisfy Partial Control, we lose the motivation for Moral Control. I have introduced Moral Control to explain why S's actions in modified Gettier-cases like Barns II, Trolley IV, and Doctor IV are only accidentally right (despite their rightness being reliable). However, if S controls some right-making features in these cases, he also controls the rightness of his actions. But then, Moral Control can't explain why S's actions appear to be only accidentally right. Partial Control is too weak.

Third, and finally, we could weaken Strong Control by combining Indirect Control and Partial Control. On this account, it would be sufficient to have *indirect* control over *some* right-making features—see Indirect Partial Control below. Indirect Partial Control fails for the same reasons as Partial Control. It cannot account for the lack of control in the modified Gettier-cases. We once again lose the motivation for Moral Control. Put simply, Indirect Partial Control is too weak because Partial Control is too weak.

Indirect Partial Control

S has control over the rightness of action φ if, and only if, S directly or indirectly controls some right-making features of φ.

I have suggested that all three straightforward weakenings of Strong Control fail. To be sure, one might defend any one of the weakenings by tweaking its formulation or application to individual cases. However, I think such efforts would be misguided. All three weakenings are in an important sense just variations on Strong Control. They maintain the core idea that control over an action's rightness *exclusively* depends on control over the right-making features. I think this core idea is fundamentally wrong.

To see why, consider HUNTER III: *S* is a completely unskilled gun fanatic. He aimlessly runs through hunting area *A* and shoots at all animals he sees. One day, *S* shoots a wolf. As it turns out, his action was (objectively) right because there were too many wolfs compared to the number of deer. However, *S* is completely unaware of the right-making features of his action. In particular, *S* doesn't know the number of wolfs and deer in *A*, and he is entirely ignorant of the relevant ecological interdependencies.

If contrasted with HUNTER III (rather than HUNTER II) it seems that *S* controls the rightness of his action in HUNTER I. Although *S* does not control the right-making features of his action, he is fully aware of the features which make his action right (in sharp contrast to HUNTER III). *S* knows the exact number of wolfs and deer in *A*, and he can act accordingly. We might say that *S* controls his action in light of its moral status. Based on these considerations, I suggest the following weakening of Strong Control:

Weak Control

S has control over the rightness of an action φ if, and only if,
(a) *S* controls the right-making features of the action φ,[55] *or*
(b) *S* controls the action φ in light of its moral status.

Before moving on, I briefly clarify what it takes to control an action in light of its moral status. The core idea is this: An agent controls the action in light of its moral status if he's responsive to its moral status. In other words, the agent controls whether or not to perform an action in response to this action being right or wrong. To be responsive to the moral status of an action, the agent (not the action) must be responsive to the right-making features.[56] If he isn't responsive to the right-making features, he violates condition (b).

Weak Control has three advantages. First, it is intuitive. Roughly speaking, it's intuitive to think that we control whether or not an action is right either by controlling the rightness or by controlling the action. More precisely, if the action is fixed, the agent controls its rightness by controlling the right-making features; this thought leads to condition (a). Conversely, if we fix the right-making features, the agent controls the action's rightness by controlling the action in light of its moral status; this suggests condition (b).

---

55 My argument in this section works even if we condition (a) is fully rejected.

56 Assuming that condition (a) is always violated, I thus hold that responsiveness to the right-making features is a necessary condition on moral worth (via Moral Incompatibility, Moral Control, and Weak Control). I thus ultimately agree with Arpaly, who claims that an action has moral worth only if the agent is "responsive[...] to the action's right-making features" (2002: 226).

Second, Weak Control is weak enough to solve the problem of moral worth without rejecting Moral Control. While most (or all) actions violate condition (a) (in the strong reading), it's easy to satisfy condition (b). We are generally able to control an action in light of its moral status. In ordinary circumstances, we are responsive to the features which make our actions right. Weak Control thus implies that we generally have control over the rightness of our actions, thereby avoiding the skeptical implications of Strong Control.

Third, and finally, Weak Control is not too weak. *S*'s actions in the modified Gettier-cases violate both conditions of Weak Control. Let me demonstrate. In BEGGARS II, *S* has no (direct) control over (all) right-making features of his donation to the person on slot s_3. To give an example, *S* does not control the fact that the person on slot s_3 is a real beggar rather than a fake one. The absence of the fake beggar is mainly ensured by the police officer. *S*'s donation thus violates condition (a) of Weak Control.

S also doesn't control the donation in light of its moral status. In the environment of fake-beggar-county, it seems that *S* is not responsive to the right-making features of his action. More specifically, it seems plausible that *S* is not responsive to the fact that the person on slot s_3 is a real beggar rather than a fake one. Remember that the fake-beggars are virtually indistinguishable from the real beggar. On the basis of these considerations, we can therefore conclude that *S*'s donation also violates condition (b) of Weak Control.

Next, let us consider Trolley IV: S has no (direct) control over (all) right-making features of his action. As a reminder, there are five people on track A and 1000 people on track B. S pulls the lever, diverting the trolley to track B. A rock rolls on track B, stops the trolley, and saves the 1000 people. (Superman would have saved the people had there been no rock.) While the rock makes S's action right, S has no plausible control over its presence. That's why S's lever-pulling violates condition (a) of Weak Control.

S also doesn't control the lever-pulling in light of its moral status. In the environment described by Trolley IV, S is not responsive to the right-making features of his action. S is not aware of the events happening many miles down track B. In particular, S doesn't know that a rock rolls on track B, stopping the trolley before it kills 1000 people. (S also doesn't know that Superman would have stopped the trolley had there been no rock.) For these reason, S's lever-pulling also violates condition (b) of Weak Control.

Finally, let us turn to Doctor IV: S has no (direct) control over (all) right-making features of his action. S prescribes medication M_1 to a patient with disease D_2. After taking the medication, the patient drinks lemon water. The lemon's acid chemically reacts with M_1, thereby curing the patient. (In nearby worlds, all water is acidic.) S's action is right because the patient drank the lemon water. However, S has no direct control over the patient's drinking habits. And so S's action violates condition (a) of Weak Control.

What is more, S also has no control over the prescription in light of its moral status. In the environment described by Doctor IV, S is not responsive to the right-making features of his action. To be more specific, S is ignorant of the patient's medical condition; S doesn't know that the patient actually has disease D_2; S is not aware of the patient's drinking habits; and S doesn't know that medication M_1 chemically reacts with the acid from the lemon. S's prescription thus violates condition (b) of Weak Control.

I have argued that S's actions in the three modified Gettier-cases (Beggars II, Trolley IV, and Doctor IV) violate both conditions of Weak Control. It follows that S doesn't control the rightness of his actions. This result is an advantage of Weak Control because it retains the motivation of Moral Control. If S doesn't control the rightness of his actions, Moral Control can explain why S's actions are merely accidentally right (despite their rightness being reliable). For this reason, I have introduced Moral Control

In summary, Weak Control has three advantages. First, it provides an intuitive clarification of what it is to control the rightness of an action. Second, it avoids the skeptical results of Strong Control. We generally have control over the rightness of our actions. Weak Control therefore defends Moral Control by solving the problem of moral worth. Third, Weak Control isn't too weak. It covers the modified moral Gettier-cases (Beggar II, Trolley IV, and Doctor IV), thus retaining the motivation for Moral Control.

My focus in this section has been on the moral domain—mainly because there is no equivalent discussion of control in the epistemic literature. However, before moving on, I should note that my considerations transfer to the epistemic domain. To begin with, we can specify the notion of epistemic control in analogy to Weak Control: *S* controls the truth of his belief that *p* if, and only if, (a) *S* controls the true-making features of his belief that *p*, or (b) *S* controls his belief that *p* in light of its truth-value.

Besides being intuitive, this specification has two advantages. First, it solves the problem of epistemic luck without having to reject Epistemic Control. While most (or all) of our beliefs violate condition (a) (see page 81), many beliefs satisfy condition (b). We are generally able to control our beliefs in light of their truth-value. In normal circumstances, we are responsive to the true-making features of our beliefs. The epistemic equivalent of Weak Control thus avoids the skeptical conclusion of the problem of epistemic luck.

Second, we can still explain why *S*'s beliefs in the modified Gettier-cases (Barns II, Clock II, and Sheep II) are only accidentally true (despite being reliable). *S* clearly doesn't control the true-making features of his beliefs (e.g., it being 3:00 pm). Arguably, *S* also isn't responsive to the true-making features and thus doesn't control his beliefs in light of their truth-values. For both reasons, *S* has no control over the truth of his beliefs. According to Epistemic Control, *S* beliefs are therefore merely accidentally true.

3.4 Restriction

I have introduced Moral Control as a necessary condition on non-accidental rightness. In this section, I explain why Moral Control isn't a sufficient condition (neither individually nor in combination with Epistemic Reliability). But first, note that *S* lacks control over the rightness of his actions not just in the *modified* Gettier-cases but also in the *original* Gettier-cases (Beggars, Trolley III, and Doctor III). In fact, the original cases violate Weak Control for the same reasons as the modified cases. A brief overview:

In Beggars, *S* doesn't (directly) control (all) right-making features of his donation to the person on slot s_3. Above all, *S* doesn't control the fact that there is a real rather than a fake beggar on slot s_3. *S* also doesn't control the donation in light of its moral status because he isn't responsive to its right-making features. In particular, he is not responsive to the fact that there is a real rather than a fake beggar on slot s_3. As a consequence, *S* doesn't control the rightness of his donation to the person on slot s_3 (cf. page 88).

In Trolley III, *S* doesn't have (direct) control over (all) right-making features of his lever-pulling. More specifically, *S* doesn't control the rock which, by rolling on track *B*, stops the trolley and saves the 1000 people. In addition, *S* doesn't control his action in light of

its moral status because he is not responsive to the rock rolling on track *B*. The rock (and the 1000 people) are many miles down the track, so that *S* can't see them. For both reasons, *S* doesn't control the rightness of his lever-pulling (cf. page 89).

In Doctor III, *S* once again has no (direct) control over (all) right-making features of his prescription. In particular, *S* does not control the patient's drinking habits. What is more, *S* does not control his action in light of its moral status because he is not responsive to the right-making features. For example, *S* doesn't know that the patient drinks lemon water and that the acid from the lemon chemically reacts with medication M_1. That is why *S* has no control over the rightness of his prescription (cf. pages 89f.).

In sum, then, *S* lacks control over the rightness of his actions in the modified as well as original Gettier-cases (the reasons being the same). Moral Control therefore covers not only the modified but also the original Gettier-cases. *S*'s actions in Beggars, Trolley III, and Doctor III are merely accidentally right because *S* doesn't control their rightness. At this point, one might be inclined to think that Moral Control is, in fact, sufficient for non-accidental rightness, not just necessary. It seems to cover all cases.

However, Moral Control doesn't cover all cases and therefore isn't sufficient. In particular, it doesn't cover the three cases from the introduction (see pages 1f.). In Shopkeeper, *S* treats his customers fairly to maximise his long-term profit. *S*'s action is only acciden-

tally right. Yet it seems that *S* has control over the rightness of his action. While he might not control its right-making features, we can assume that *S* is sufficiently responsive to the features which make his action right—though he doesn't act on these features.

Similar considerations apply to the Brother case. *S* sees his brother drowning in a pond. The brother owes *S* a lot of money, and so *S* saves him only to get his money back. *S*'s action is only accidentally right, but *S* has control over its rightness. To be sure, *S* might not control all right-making features of the action. Yet it seems that *S* is responsive to the right-making features, or so we can assume. For instance, he is aware that his brother is drowning and needs help. But again, *S* doesn't act on these features.

Finally, consider Politician: *S* is a career politician, serving meals in a soup kitchen only to win the upcoming election. Although *S*'s action is only accidentally right, he has control over its rightness. Again, *S* might not control all right-making features (though as a politician he has some influence on the people's poverty). But we can assume that *S* is responsive to the right-making features of his action. He is aware of the poverty, and the people's needs. However, as before, he doesn't act on these features.

Moral Control is not sufficient for non-accidental rightness because it doesn't cover Shopkeeper, Brother, and Politician. However, as I have previously argued (see pages 5, 20), Moral Reliability covers Shopkeeper, Brother, and Politician: *S*'s actions are only

accidentally right because their rightness isn't reliable. Maybe then Moral Control and Moral Reliability aren't individually but jointly sufficient for non-accidental rightness. Together they are seemingly able to cover all cases of non-accidental rightness.

I argue Moral Control and Moral Reliability don't cover all cases and thus are not sufficient for non-accidental rightness. Some actions are only accidentally right, but their rightness is both controlled and reliable. Remember that I have previously modified the moral Gettier-cases so that *S*'s actions are accidentally right and reliable (see §§ 2.3 and 2.5). Shopkeeper, Brother, and Politician can be modified to the same effect. *S*'s actions are then accidentally right but satisfy Moral Reliability (in addition to Moral Control).

Consider Shopkeeper II: *S* is still a prudent shopkeeper who treats his customers fairly to maximise his long-term profit. *S* has a fixed customer base and maintains a good reputation to stay in business. But this time, we assume that the invisible hand ensures that self-interest and morality necessarily align. To be sure, *S*'s conduct still is right only by accident. However, now there are no nearby worlds in which *S*'s self-interested action won't match the moral demands. As a result, *S*'s fair treatment is reliably right.[57]

[57] Arpaly states the same case in a different context: "We can imagine a world in which some invisible hand makes the profit motive reliably produce morally right actions, place Kant's grocer in that world, and still not free ourselves from the sense that there is something accidental in his acting well" (2002: 225). See also Herman (1981: 366), Benson (1987: 367), Arpaly (2003), (2015: 88), Smilansky (2005: 492f.), Markovits (2010: 211n.), and Lord (2015).

Next, consider Brother II: Again, *S* saves his drowning brother only to get back the money his brother still owes him. Now let us assume that *S*'s brother is notoriously broke; he just cannot deal with money. The brother also doesn't know anyone except *S* (he has no other relatives or friends). He thus asks *S* for money in most nearby worlds. These considerations don't change the fact that *S*'s action is only accidentally right. But this time, *S* saves his brother in most nearby worlds. His action is reliably right.

And finally, Politician II: As in the original case, *S* is a politician who serves meals in a soup kitchen to win the election. Now let us change the details of the political system. Suppose the people could hold a vote of no-confidence at any point in time. In this scenario, *S* serves meals in soup kitchens not only during election time, for he must maintain his caring image at all times. Still, his action is only accidentally right. However, this time, he will serve the meals in most nearby worlds. *S*'s action is reliably right.

S's actions in Shopkeeper II, Brother II, and Politician II are accidentally right, but their rightness is reliable. Note that nothing depends on the details of the cases. As a last resort, we can always introduce a benevolent demon, ensuring the actions being reliably right. And as in the original cases (see pages 92f.), *S* also controls the rightness of his actions. Summarised in one sentence: *S*'s actions in Shopkeeper II, Brother II, and Politician II are accidentally right, but their rightness is both controlled and reliable.

In this section, I have argued that Moral Control is not sufficient for non-accidental rightness—neither individually nor in combination with Moral Reliability. Before I conclude this thesis, I should briefly note that my discussion of Moral Control generalises to the epistemic domain. Epistemic Control isn't individually sufficient for non-accidental truth. We can imagine a case in which the agent is generally responsive to the true-making features but forms his belief on the basis of ulterior considerations.

Epistemic Control and Epistemic Reliability are not jointly sufficient for non-accidental truth. As in the moral domain, we could modify a controlled but accidentally true belief to make its truth reliable. For instance, *S* is responsive to the true-making features (controlled); his belief is based on ulterior consideration (accidental); but a demon ensures the truth of *S*'s belief (reliable).[58] *S*'s belief is accidentally true but satisfies Epistemic Control and Epistemic Reliability. I leave it with these sketchy remarks—my focus is on the moral domain.

[58] The benevolent demon has to intervene a lot if we were to specify reliable truth along the lines of Epistemic Safety *Refined*. Not only would the demon have to ensure that *S*'s believe in *p* isn't false in most (or all) nearby worlds, the demon would also have to ensure that *S*'s ulterior method is reliable (see page 51). For this purpose, the demon would have to make *S*'s beliefs in propositions other than *p* true as long as *S* uses the same ulterior method. While being more demanding, such a benevolent demon is surely possible.

4 Conclusion

I have begun my thesis with Moral Incompatibility: If S's action φ has moral worth, then S's action φ is non-accidentally right. It was my goal to clarify the relevant notion of non-accidental rightness. To this end, I have drawn on an analogy between moral worth and knowledge. Both moral worth and knowledge exclude luck. I have used the extensive literature on non-accidental truth to clarify the relevant notion of non-accidental rightness. With this methodology, I have identified two necessary conditions:

Moral Reliability

S's action φ is non-accidentally right only if
S's action φ is reliably right.

Moral Control

S's action φ is non-accidentally right only if S has control over the rightness of his action φ.

The argument of my thesis has two main parts. In the first part (§ 2), I have motivated the modal account of non-accidental rightness. The modal account is both intuitive and covers many cases—including the original moral Gettier-cases. However, both specifications of the

modal account in terms of tracking and safety fail to account for the modified moral Gettier-cases. I have therefore restricted the modal account. According to Moral Reliability, reliable rightness is only a necessary but not sufficient condition for non-accidental rightness.

In the second part (§ 3), I have suggested that the modal condition must be supplemented with Moral Control to cover the modified moral Gettier-cases. An action is non-accidentally right only if the agent controls the action's rightness. I have defended this control condition against a variation of the problem of moral luck, arguing that Moral Control does not have skeptical consequences. Lastly, I have noted that Moral Control is not sufficient for non-accidental rightness—neither individually nor with Moral Reliability.

As I have stated in the introduction (see pages 11f.), my thesis has three essential limitations. In the following, I briefly sketch some strategies for how one might overcome these limitations. First, I have not developed a sufficient condition for non-accidental rightness. On the contrary, I have repeatedly claimed and argued that Moral Reliability and Moral Control are neither individually nor jointly sufficient for non-accidental rightness (§§ 2.6 and 3.4). But what has to be added, in order to generate a sufficient account?

To answer this question, it's instructive to revisit the counterexamples to the joint sufficiency of Moral Reliability and Moral Control. In Shopkeeper II, Brother II, and Politician II, *S*'s actions are accidentally right but their rightness is both reliable and controlled.

What strikes me about these cases is that *S has* control over the actions in light of their moral status—for *S* is generally responsive to the right-making features. However, *S* does not *exercise* that control. *S* does not act on the basis of the right-making features.

One might therefore cover cases like Shopkeeper II, Brother II, and Politician II by strengthening the control condition: *S*'s action φ is non-accidentally right if, and only if, *S exercises* his control over the rightness of φ. (I drop the modal condition for now.) Even if this strengthened control condition covers Shopkeeper II, Brother II, and Politician II, one might be concerned that it fails to account for cases in which the agent exercises his control in some deviant way. In such cases, *S*'s actions might still be only accidentally right.

Here is my suggestion how we can overcome this further difficulty: In those cases where the agent exercises his control over the action's rightness in some deviant way, it seems that the action's rightness cannot be explained by the fact that the agent exercised his control. In some relevant sense, the action is not right *because* the agent has exercised his control over its rightness. On the basis of these considerations, one might advocate for a further strengthening of the control condition. I therefore suggest:

Strong Moral Control

S's action φ is non-accidentally right if, and only if, φ is right because of *S* exercising his control over the action's rightness.

Strong Moral Control resembles virtue-theoretic accounts of knowledge. Virtue epistemologists commonly argue that a belief amounts to knowledge, if and only if, its truth can be explained by the agent's exercise of his cognitive abilities: "[K]nowledge is cognitive success that is because of the exercise of the relevant cognitive abilities" (Pritchard 2012: 265).[59] If we identify epistemic virtues with epistemic control rather than cognitive abilities (see pages 67f.), we can also state a virtue-theoretic account of non-accidental truth:

Strong Epistemic Control

S's belief that *p* is non-accidentally true if, and only if, *p* is true because of *S* exercising his control over the belief's truth.

What's with the modal conditions? In the epistemic literature, it has been argued that all virtuous beliefs are reliably true. In the pointed words of Greco: "[V]irtuous belief entails [...] safety" (forthcoming).[60] If true—and this is a big *if*—, one might also think that an action is reliably right when it is virtuous, i.e., when its rightness is explained by the agent's exercise of his control over the action's rightness. Strong Moral Control might thus entail Moral Reliability. If true, we would get a unified account of non-accidental rightness.

[59] Greco et al. similarly state: "[T]o know is to believe the truth because of your intellectual virtue" (2018). See also Sosa (1991: 277), Zagzebski (1996: 271f.), Lehrer (2000: 223), Riggs (2002: 93f.), Greco (2003: 111), and Turri (2011).

[60] For a further discussion of the purported entailment, see (among others) Pritchard (2005: 130), (2009a: 41f.), (2012: 264), Sosa (2007), (2015: 224), Greco (2010: 77), Greco et al. (2018), Carter et al. (forthcoming), and Das (n.d.).

Second, even a complete account of non-accidental rightness might not be sufficient for moral worth. This is because moral worth might require more than just non-accidental rightness. But what in addition to non-accidental rightness could be required for moral worth? To answer this question, it is again instructive to first examine the epistemic literature. There it has been argued that "eliminat[ing] [...] epistemic luck will [not] be enough to offer a complete analysis of knowledge" (Pritchard 2015: 96)

More specifically, internalists have argued that an anti-luck condition is not sufficient for knowledge because it "naturally leads to an externalist epistemology—i.e., an epistemology which does not demand of knowers that they are able to offer an internalist justification for their beliefs" (Pritchard 2007: 285).[61] In other words, the internalist claims that knowledge requires internal justification (i.e., our beliefs must have reflectively accessible grounds). Yet a belief lacking internal justification can be non-accidentally true.

As an example for an unjustified but non-accidentally true belief, consider the famous Chicken-Sexer case:[62] *S* is a chicken-sexer who reliably detects the sex of chickens based on some sub-personal olfactory abilities. Being completely unaware of these abilities, *S*

[61] For the purpose of this thesis, I assume that internalism states an internal justification condition on knowledge. There are other definitions of epistemic internalism. See Sosa (1999a: 373), Latus (2000: 162), and Pritchard (2005: 42).

[62] Pritchard argues that the "chicken-sexer['s] [...] beliefs [...] fail to meet an internalist justification condition" (2005: 74). For an extended formulation and discussion of the Chicken-Sexer case, see Pritchard (2006: 15), (2007), (2009b: 445f.), (2012: 262), Becker (2008b: 355), as well as Breyer (2010: 135).

cannot cite any grounds for his beliefs. At best, he can give some spurious explanations. While externalists think that *S* has knowledge, "[i]nternalists characteristically deny knowledge" because *S* "lacks good reflectively accessible grounds" (Pritchard 2009b: 456).

But *S*'s beliefs in Chicken-Sexer are non-accidentally true; even under Strong Epistemic Control. We can assume that *S* is subconsciously responsive to the true-making features of his beliefs. He thus controls his beliefs in light of their truth-value. It also seems that his beliefs being true can be explained by the exercise of his control. In conclusion, then, if we think that internalism about knowledge is true—again, this is a big *if*—non-accidental truth is not sufficient for knowledge, but we would have to add a justification condition.

Analogous considerations apply to the moral domain. In the literature on moral worth, one can find a moral analogue to the Chicken-Sexer (though it is never classified as such), namely the case of Huckleberry Finn: Huck does the right-thing and frees his friend Jim from slavery. His action is based on a strong sub-personal feeling that slavery is inhumane. However, when asked, Huck would say that he did the wrong thing. Huck thinks that freeing Jim was theft, for Jim was the legal property of Miss Watson.[63]

[63] In the pointed words of Isserow: "Huck regards slavery as a form of ownership, and so he famously suffers from a guilty conscience when he [...] helps [Jim] to escape" (2018: 4). For an extended formulation and discussion of the Huckleberry Finn case, see Arpaly (2002: 230f.), (2003: 74-79), (2015: 86), Markovits (2010: 218), (2012: 293f.), Arpaly & Schroeder (2014: 178f.), Sliwa (2015: 180), (2016: 415f.), Lord (2017: 456), as well as Johnson King (n.d.).

Huck's action was non-accidentally right; even according to the high standards of Strong Moral Control. Korsgaard (1996b) suggests that moral feelings can generally be responsive to the right-making features of our actions. In this context, one might argue that Huck's feeling for the inhumanity of slavery is responsive to the right-making features of his action. Huck therefore controls the action in light of its moral status. Moreover, it seems that the exercise of this control explains why Huck did the right thing.

Mirroring epistemic internalism, it has been suggested that Huck's freeing of Jim lacks moral worth—just like the chicken-sexer lacks knowledge: "Huckleberry's action [...] is not morally praiseworthy" (Sliwa 2016: 415; see also 2015: 180). If internalism about moral worth is true, we can conclude that non-accidental rightness is not sufficient for moral worth. In analogy to epistemic internalism, one might thus further suggest that moral worth requires the agent to have reflective access to the grounds for his actions.

This being said, one might think that the whole project of finding sufficient conditions for moral worth is misguided. Knowledge-first epistemologists argue that knowledge is a primitive notion that can not be further analysed. On the contrary, other epistemic terms like 'belief' might be analysed in terms of knowledge. Likewise, one might propose a moral-worth-first ethics, according to which moral worth is a primitive unanalysable notion. On this view, the project of analysing moral worth would be misguided.

Third, my discussion in this thesis does not account for degrees of moral worth and non-accidental rightness. However, one could argue that reliable rightness and control come in degrees—surely the rightness of an action can be more or less reliable, and the agent can have more or less control. The degree of non-accidental rightness (and moral worth) might then be a function of the degrees of reliable rightness and moral control. The binary concepts can then be recovered by introducing a threshold.

In short, I have sketched three strategies for how to overcome the limitations of my thesis. I have suggested that we could strengthen the control condition to get a sufficient account of non-accidental rightness. I have then indicated that moral worth might require justification in addition to non-accidental rightness. Lastly, I have suggested that degrees of non-accidental rightness can be explained by degrees of reliability and control. To be clear, these strategies are preliminary and require extensive future work.

In closing this thesis, I highlight three avenues along which one might further develop the analogy between moral worth and knowledge. First, we could expand the analogy. In the introduction to *Knowledge and Its Limits*, Williamson writes, "desires can [...] be satisfied and beliefs true by chance" (2000: 1). If accidentally true beliefs don't amount to knowledge, what do accidentally satisfied desires fall short of? Happiness, maybe. Knowledge excludes luck; moral worth excludes luck; maybe happiness excludes luck, too.

Even if the intuition that happiness excludes luck might not be as strong as in the epistemic and moral case, it is still worth thinking about happiness in these terms. For a lot of interesting questions arise in this context: Does happiness require reliable desire satisfaction? And if so, should it be specified in terms of tracking or safety? Do we have to control the satisfaction of our desires? And if so, should we control the satisfaction-making features? Or should we control our desires in light of their satisfaction-status?

Second, we might use the analogy to explain the distinctive value of knowledge. Knowledge seems more valuable than merely true belief (Kvanvig 2003, Pritchard 2009c). But why? Maybe it's easier to explain why moral worth is more valuable than merely right action; or why happiness is more valuable than merely satisfied desire. If true, we could use the structural insights from the moral and prudential domains to make progress in explaining why knowledge appears to have a distinctive value.

Finally, the analogy to knowledge (and happiness) might be used to further motivate the topic of moral worth. While the notion of moral worth has received some attention in the last two decades (in addition to Kant's seminal discussion), the literature diminishes when compared to the centuries of work on the notion of rightness. Just imagine epistemology had done the equivalent thing and focused on truth rather than knowledge. Maybe moral worth should be to ethics, what knowledge is to epistemology.

5 Bibliography

Allison, Henry E. (2011): *Kant's Groundwork for the Metaphysics of Morals, A Commentary*, Oxford: Oxford University Press.

Archer, Alfred (2016): "Are Acts of Supererogation Always Praiseworthy?", in *Theoria*, 82, 238-255.

Arpaly, Nomy (2002): "Moral Worth," in *Journal of Philosophy*, 99, 223-245.

Arpaly, Nomy (2003): *Unprincipled Virtue, An Inquiry into Moral Agency*, Oxford: Oxford University Press.

Arpaly, Nomy (2015): "Moral Worth and Normative Ethics", in: Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 5*, Oxford: Oxford University Press: 87-105.

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

Axtell, Guy (2003): "Felix Culpa: Luck in Ethics and Epistemology", in *Metaphilosophy*, 34, 331-352.

Bader, Ralf (n.d.): "The Dignity of Humanity" (manuscript).

Baron, Marcia W. (1995): *Kantian Ethics almost without Apology*, Ithaca: Cornell University Press.

Baumann, Peter (2012): "No Luck With Knowledge? On a Dogma of Epistemology", in *Philosophy and Phenomenological Research*, 89, 523-551.

Becker, Kelly (2007): *Epistemology Modalized*, London: Routledge.

Becker, Kelly (2008a): "Defending a Sensitive Neo-Moorean Invariantism", in: Duncan Pritchard et al. (eds.), *New Waves in Epistemology*, Basingstoke: Palgrave Macmillan, 8-27.

Becker, Kelly (2008b): "Epistemic Luck and the Generality Problem", in *Philosophical Studies*, 139, 353-366.

Beddor, Bob & Pavese, Carlotta (forthcoming): "Modal Virtue Epistemology", in *Philosophy and Phenomenological Research*.

Beebe, James R. (2004): "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis", in *Noûs*, 38, 177-195.

Benson, Paul (1987): "Moral Worth", in *Philosophical Studies*, 51, 365-382.

Benzenberg, Christopher P. (n.d.): "Oops! ...I Did the Right Thing" (manuscript).

Black, Tim (2002): "A Moorean Response to Brain-in-a-Vat Scepticism", in *Australasian Journal of Philosophy*, 80, 148-163.

Black, Tim & Murphy, Peter (2007): "In Defense of Sensitivity", in *Synthese*, 154, 53-71.

Bondy, Patrick & Pritchard, Duncan (2018): "Propositional Epistemic Luck, Epistemic Risk, and Epistemic Justification", in *Synthese*, 195, 3811-3820.

Brandt, Richard B. (1959): *Ethical Theory*, *The Problems of Normative and Critical Ethics*, Englewood Cliffs: Prentice-Hall.

Breyer, Daniel (2010): "Reflective Luck and Belief Ownership", in *Acta Analytica*, 25, 133-154.

Broncano-Berrocal, Fernando (2015): "Luck as Risk and the Lack of Control Account of Luck", in *Metaphilosophy*, 46, 1-25.

Carr, Jennifer (2015): "Subjective Ought", in *Ergo*, 2, 678-710.

Carrier, Leonard S. (1971): "An Analysis of Empirical Knowledge", in *Southern Journal of Philosophy*, 9, 3-11.

Carter, Adam J. et al. (forthcoming): "Varieties of Cognitive Achievement", in *Philosophical Studies*.

Chisholm, Roderick M. (1957): *Perceiving*, *A Philosophical Study*, Ithaca: Cornell University Press.

Chisholm, Roderick M. (1989): *Theory of Knowledge*, Englewood Cliffs: Prentice Hall.

Clarke-Doane, Justin (2016): "What Is the Benacerraf Problem?", in: Fabrice Pataut (ed.), *New Perspectives on the Philosophy of Paul Benacerraf, Truth, Objects, Infinity*, Berlin: Springer, 17-43.

Coffman, E. J. (2007): "Thinking About Luck", in *Synthese*, 158, 385-398.

Coffman, E. J. (2009): "Does Luck Exlude Control?", in *Australasian Journal of Philosophy*, 87, 499-504.

Dancy, Jonathan (1985): *Introduction to Contemporary Epistemology*, Oxford: Blackwell.

Das, Nilanjan (n.d.): "Knowledge and Moral Worth" (manuscript).

de Castro, Rodrigo J. (2005): "Is Moral Worth Compatible with Cooperating Inclinations?", in *Princípios*, 12, 5-18.

DeRose, Keith (1995): "Solving the Skeptical Problem", in *Philosophical Review*, 104, 1-52.

DeRose, Keith (2002): "Sosa, Safety, Sensitivity, and Skeptical Hypotheses", in: John Greco (ed.), *Ernest Sosa, And His Critics*, Oxford: Blackwell, 22-41.

Dorsey, Dale (2012): "Objective Morality, Subjective Morality and the Explanatory Question", in *Journal of Ethics and Social Philosophy*, 6, 1-24.

Douglas, Thomas (2014a): "Enhancing Moral Conformity and Enhancing Moral Worth", in *Neuroethics*, 7, 75-91.

Douglas, Thomas (2014b): "The Relationship Between Effort and Moral Worth: Three Amendments to Sorensen's Model", in *Ethical Theory and Moral Practice*, 17, 325-334.

Dretske, Fred (1970): "Epistemic Operators", in *Journal of Philosophy*, 67, 1007-1023.

Dretske, Fred (1971): "Conclusive Reasons", in *Australasian Journal of Philosophy*, 49, 1-22.

Engel, Mylan Jr. (1992): "Is Epistemic Luck Compatible With Knowledge?", in *Southern Journal of Philosophy*, 30, 59-75.

Engel, Mylan Jr. (2019): "Epistemic Luck", in *Internet Encyclopedia of Philosophy*, https://www.iep.utm.edu/epi-luck/.

Ewing, Alfred (1948): *The Definition of Good*, London: Routledge & Kegan Paul.

Faraci, David (2019): "Groundwork for an Explanationist Account of Epistemic Coincidence", in *Philosopher's Imprint*, 19, 1-26.

Feldman, Fred (1986): *Doing the Best We Can, An Essay in Informal Deontic Logic*, Dordrecht: Reidel.

Feldman, Fred (2012): "True and Useful: On the Structure of a Two Level Normative Theory", in *Utilitas*, 24, 151-171.

Forbes, Graeme (1984): "Nozick on Scepticism", in *Philosophical Quarterly*, 34, 43-52.

Freitag, Wolfgang (2014): "Safety, Sensitivity and 'Distant' Epistemic Luck", in *Theoria*, 80, 44-61.

Gaultier, Benoit (2014): "Achievements, Safety and Environmental Epistemic Luck", in *Dialectica*, 68, 477-497.

Gert, Joshua (2012): "Moral Worth, Supererogation, and the Justifying/Requiring Distinction", in *Philosophical Review*, 121, 611-618.

Goldman, Alvin (1976) "Discrimination and Perceptual Knowledge", in *Journal of Philosophy*, 78, 771-791.

Goldman, Alvin (1983): "Review of Philosophical Explanations by Robert Nozick", in *Philosophical Review*, 92, 81-88.

Goldman, Alvin & Beddor, Bob (2016): "Reliabilist Epistemology", in *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/win2016/entries/reliabilism/.

Graham, Peter A. (2010): "In Defense of Objectivism about Moral Obligation", in *Ethics*, 121, 88-115.

Greco, John (1995): "A Second Paradox Concerning Responsibility and Luck", in *Metaphilosophy*, 26, 81-96.

Greco, John (1999): "Agent Reliabilism", in *Philosophical Perspectives*, 13, 273-296.

Greco, John (2003): "Knowledge as Credit for True Belief", in: Michael DePaul & Linda Zagzebski (eds.), *Intellectual Virtues, Perspectives from Ethics and Epistemology*, Oxford: Oxford University Press, 111-134.

Greco, John (2007): "The Nature of Ability and the Purpose of Knowledge", in *Philosophical Issues*, 17, 57-68.

Greco, John (2010): *Achieving Knowledge, A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge: Cambridge University Press.

Greco, John (2012): "A (Different) Virtue Epistemology", in *Philosophy and Phenomenological Research*, 85, 1-26.

Greco, John (forthcoming): "Safety in Sosa", in *Synthese*.

Greco, John et al. (2018): "Virtue Epistemology", in *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/sum2018/entries/epistemology-virtue/.

Greene, Preston (forthcoming): "When Is a Belief True Because of Luck?", in *Philosophical Quarterly*.

Guyer, Paul (2007): *Kant's Groundwork for the Metaphysics of Morals, A Reader's Guide*, London: Continuum.

Guyer, Paul (2018): "Moral Worth and Moral Motivation: Kant's Real View", in: Sally Sedgwick & Dina Emundts (eds.), *Begehren/Desire*, Berlin: De Gruyter, 19-38.

Hales, Steven D. (2016): "Why Every Theory of Luck is Wrong", in *Noûs*, 50, 490-508.

Hansson, Sven O. (2010): "Objective or Subjective 'Ought'?", in *Utilitas*, 22, 33-35.

Harper William (1996): "Knowledge and Luck", in *Southern Journal of Philosophy*, 34, 273-283.

Henson, Richard G. (1979): "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action", in *Philosophical Review*, 88, 39-54.

Herman, Barbara (1981): "On the Value of Acting from the Motive of Duty", in *Philosophical Review*, 90, 359-382.

Hetherington, Stephen (2013): "There Can Be Lucky Knowledge", in: Matthias Steup et al. (eds.), *Contemporary Debates in Epistemology*, Oxford: Blackwell, 164-176.

Hiller, Avram & Neta, Ram (2007): "Safety and Epistemic Luck", in *Synthese*, 158, 303-313.

Hills, Alison (2009): "Moral Testimony and Moral Epistemology", in *Ethics*, 120, 94-127.

Hills, Alison (2011): "Moral Epistemology", in: Michael Brady (ed.), *New Waves in Metaphysics*, New York: Palgrave Macmillican, 249-263.

Hills, Alison (2012): "Comment on Karen Jones and Francois Schroeter", in *Analyse & Kritik*, 231-236.

Hills, Alison (2013): "Moral Testimony", in *Philosophy Compass*, 8, 552-559.

Hirevlä, Jaakko (2019): "Gloabel Safty", in *Synthese*, 196, 1167-1186.

Ho, Tsung-Hsing (forthcoming): "Lucky Achievements: Virtue Epistemology on the Value of Knowledge", in *Ratio*.

Hurley, Susan L. (2003): *Justice, Luck, and Knowledge*, Cambridge: Harvard University Press.

Ichikawa, Jonathan J. & Steup, Matthias (2018): "The Analysis of Knowledge", in *Stanford Encyclopaedia of Philosophy*, https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/.

Isserow, Jessica (2018): "Moral Worth and Doing the Right Thing by Accident", in *Australasian Journal of Philosophy*, 97, 1-14.

Jackson, Frank (1986): "A Probabilistic Approach to Moral Responsibility", in: Ruth Barcan Marcus et al. (eds.), *Logic, Methodology, and Philosophy of Science VII*, North Holland: Elsevier Science Publishers, 351-365.

Johnson King, Zoë A. (2018a): "Accidentally Doing the Right Thing", in *Philosophy and Phenomenological Research*, Early View, 1-21.

Johnson King, Zoë A. (2018b): *Trying to Act Rightly*, PhD, Ann Arbor: The University of Michigan.

Johnson King, Zoë A. (n.d.): "Accidentally Doing the Right Thing" (manuscript).

Kallestrup, Jesper & Pritchard, Duncan (2011): "Virtue Epistemology and Epistemic Twin Earth", in *European Journal of Philosophy*, 22, 335-357.

Kant, Immanuel (1900–): *Kant's Gesammelte Schriften*, Königlich-Preußische Akademie der Wissenschaften (ed.), Berlin: De Gruyter.

Kelp, Christoph (2009a): "Knowledge and Safety", in *Journal of Philosophical Research*, 34, 21-31.

Kelp, Christoph (2009b): "Pritchard on Knowledge, Safety, and Cognitive Achievements", in *Journal of Philosophy*, 34, 51-53.

Korsgaard, Christine M. (1996a): *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.

Korsgaard, Christine M. (1996b). "From Duty and For The Sake of The Noble", in: Stephen Engstrom & Jennifer Whiting (eds.), *Aristotle, Kant, and the Stoics*, *Rethinking Happiness and Duty*, Cambridge: Cambridge University Press, 203-236.

Kripke, Sauk (2011): "Nozick on Knowledge", in: idem (ed.), *Philosophical Troubles*, *Collected Papers*, *Volume 1*, Oxford: Oxford University Press: 162-224.

Kvanvig, Jonathan L. (2003): *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.

Kvanvig, Jonathan L. (2004): "Nozickian Epistemology and the Value of Knowledge", in *Philosophical Issues*, 14, 201-218.

Lackey, Jennifer (2006): "Review: Pritchard's Epistemic Luck", in *Philosophical Quarterly*, 56, 284-289.

Lackey, Jennifer (2008): "What Luck is Not", in *Australasian Journal of Philosophy*, 86, 255-267.

Latus, Andrew (2000): "Moral and Epistemic Luck", in *Journal of Philosophical Research*, 35, 149-172.

Latus, Andrew (2019): "Moral Luck", in *Internet Encyclopaedia of Philosophy*, https://www.iep.utm.edu/moralluc/.

Lehrer, Keith (1971): "How Reasons Give Us Knowledge: Or the Case of the Gypsy Lawyer", in *Journal of Philosophy*, 68, 78-82.

Lehrer, Keith (1990): *Theory of Knowledge*, Boulder: Westview Press.

Lehrer, Keith (2000): *Theory of Knowledge*, 2nd edition, Boulder: Westview Press.

Lewis, David (1973): *Counterfactuals*, Oxford: Blackwell.

Lewis, David (1986): *On the Plurality of Worlds*, Cambridge: Cambridge University Press.

Lord, Errol (2015): "Acting for the right Reasons, Abilities, and Obligation", in: Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, *Volume 10*, Oxford: Oxford University Press: 26-52.

Lord, Errol (2017): "On The Intellectual Conditions for Responsibility: Acting for the Right Reasons, Conceptualization, and Credit", in *Philosophy and Phenomenological Research*, 95, 436-464.

Luper, Steven (1984): "The Epistemic Predicament", in *Australasian Journal of Philosophy*, 62, 26-50.

Luper, Steven (1987): *The Possibility of Knowledge*, *Nozick and His Critics*, Totowa: Rowman & Littlefield.

Luper, Steven (2012): "False Negatives", in: Kelly Becker & Tim Black (eds.), *The Sensitivity Principle in Epistemology*, Cambridge: Cambridge University Press, 207-226.

Mackie, John L. (1969): "The Possibility of Innate Knowledge", in *Proceedings of the Aristotelian Society*, 70, 245-257.

Markovits, Julia (2010): "Acting for the Right Reasons", in *Philosophical Review*, 119, 201-242.

Markovits, Julia (2012): "Saints, Heroes, Sages, and Villains", in *Philosophical Studies*, 158, 289-311.

Massoud, Amy (2016): "Moral Worth and Supererogation", in *Ethics*, 126, 690-710.

McGrath, Sarah (2011): "Skepticism About Moral Expertise as a Puzzle For Moral Realism", in *Journal of Philosophy*, 108, 111-137.

McNamara, Paul (2011): "Supererogation, Inside and Out: Toward an Adequate Scheme for Common-Sense Morality", in: Mark Timmons (ed.), *Oxford Studies in Normative Ethics, Volume 1*, Oxford: Oxford University Press, 202-235.

Miscevic, Nenad (2007): "Armchair Luck", in *Acta Analytica*, 22, 1-22.

Morillo, Carolyn R. (1984): "Epistemic Luck, Naturalistsic Epistemology, and the Ecology of Knowledge", in *Philosophical Studies*, 46, 109-129.

Murphy, Peter (2005): "Closure Failures for Safety", in *Philosophia*, 33, 331-334.

Nagel, Thomas (1979): "Moral Luck", in: idem (ed.), *Mortal Questions*, Cambridge: Cambridge University Press, 24-38.

Nelkin, Dana K. (2016): "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness", in *Noûs*, 50, 356-378.

Nelkin, Dana K. (2019): "Moral Luck", in *Stanford Encyclopaedia of Philosophy*, https://plato.stanford.edu/archives/sum2019/entries/moral-luck/.

Nickel, Philip (2001): "Moral Testemony and Its Asymmetry", in *Ethical Testimony and Moral Practice*, 4, 253-266.

Nozick, Robert (1981): *Philosophical Explanations*, Cambridge: Harvard University Press.

Oddie, Graham & Menzies, Peter (1992): "An Objectivist's Guide to Subjective Value", in *Ethics*, 102, 512-533.

Olsen, Kristian (2017): "A Defense of the Objective/Subjective Moral Ought Distinction", in *Journal of Ethics*, 21, 351-373.

Parfit, Derek (2011): *On What Matters, First Volume*, Oxford: Oxford University Press.

Pollock, John (1979): "A Plethora of Epistemological Theories", in: George S. Pappas (ed.), *Knowledge and Justification*, *New Studies in Epistemology*, Dordrecht: Reidel, 93-113.

Portmore, Douglas W. (2011): *Commonsense Consequentialism*, *Wherein Morality Meets Rationality*, Oxford: Oxford University Press.

Pritchard, Duncan (2002): "Resurrecting the Moorean Response to the Sceptic", in *International Journal of Philosophical Studies*, 10, 283-307.

Pritchard, Duncan (2005): *Epistemic Luck*, Oxford: Oxford University Press.

Pritchard, Duncan (2006): "Moral and Epistemic Luck", in *Metaphilosophy*, 37, 1-25.

Pritchard, Duncan (2007): "Anti-Luck Epistemology", in *Synthese*, 158, 277-297.

Pritchard, Duncan (2009a): "Safety Based Epistemology: Wither Now?", in *Journal of Philosophical Research*, 34, 33-45.

Pritchard, Duncan (2009b): "Sensitivity, Safety, and Antiluck Epistemology", in: John Greco (ed.), *The Oxford Handbook of Skepticism*, Oxford: Oxford University Press, 437-455.

Pritchard, Duncan (2009c): "The Value of Knowledge", in *Harvard Review of Philosophy*, 16, 86-103.

Pritchard, Duncan (2012): "Anti-Luck Virtue Epistemology", in *Journal of Philosophy*, 109, 247-279.

Pritchard, Duncan (2014): "The Modal Account of Luck", in *Metaphilosophy*, 45, 594-619.

Pritchard, Duncan (2015): "Anti-luck Epistemology and the Gettier Problem", in *Philosophical Studies*, 172, 93-111.

Richards, Norvin (1986): "Luck and Desert", in *Mind*, 65, 198-209.

Riggs, Wayne (2002): "Reliability and the Value of Knowledge", in *Philosophy and Phenomenological Research*, 64, 79-96.

Riggs, Wayne (2007): "Why Epistemologists Are so down on Their Luck", in *Synthese*, 329-344.

Riggs, Wayne (2009a): "Luck, Knowledge, and Control", in: Duncan Pritchard et al. (eds.), *Epistemic Value*, Oxford: Oxford University Press, 204-222.

Riggs, Wayne (2009b): "Two Problems of Easy Credit", in *Synthese*, 169, 201-216.

Riggs, Wayne (2014): "Luck, Knowledge, and 'Mere' Coincidence", in *Metaphilosophy*, 45, 628-639.

Roland, Jeffrey & Cogburn, Jon (2011): "Anti-Luck Epistemologies and Necessary Truths", in *Philosophia*, 39, 547-561.

Russell, Bertrand (1912): *The Problems of Philosophy*, London: Williams & Norgate.

Russell, Bertrand (1948): *Human Knowledge, Its Scope and Limits*, New York: Allen & Unwin.

Russell, Bertrand (1966): "The Elements of Ethics", in: idem (ed.), *Philosophical Essays*, New York: Simon & Schuster, 13-59.

Sainsbury, Richard M. (1997): "Easy Possibilities", in *Philosophy and Phenomenological Research*, 57, 907-919.

Schafer, Karl (2014): "Knowledge and Two Forms of Non-Accidental Truth", in *Philosophy and Phenomenological Research*, 89, 373-393.

Setiya, Kieran (2012): *Knowing Right From Wrong*, Oxford: Oxford University Press.

Shatz, David (1987): "Nozick's Conception of Skepticism", in: Steven Luper (ed.), *Thes Possibility of Knowledge, Nozick and His Critics*, Totowa: Rowman & Littlefield, 242-266.

Shope, Robert K. (1983): *The Analysis of Knowledge, A Decade of Research*, Princeton: Princeton University Press.

Sliwa, Paulina (2012): "In Defense of Moral Testimony", in *Philosophical Studies*, 158, 175-195.

Sliwa, Paulina (2015): "Praise without Perfection: A Dilemma for Right-Making Reasons", in *American Philosophical Quarterly*, 52, 171-185.

Sliwa, Paulina (2016): "Moral Worth and Moral Knowledge", in *Philosophy and Phenomenological Research*, 93, 393-418.

Smilansky, Saul (2005): "The Paradoxical Relationship between Morality and Moral Worth", in *Metaphilosophy*, 36, 490-500.

Smith, Holly M. (2010): "Subjective Rightness", in *Social Philosophy and Policy*, 27, 64-110.

Sorensen, Kelly (2004): "The Paradox of Moral Worth", in *Journal of Philosophy*, 101, 465-483.

Sorensen, Kelly (2010): "Effort and Moral Worth", in *Ethical Theory and Moral Practice*, 13, 89-109.

Sorensen, Kelly (2014): "Counterfactual Situations and Moral Worth", in *Journal of Moral Philosophy*, 11, 294-319.

Sosa, Ernest (1974): "How Do We Know?", in *American Philosophical Quarterly*, 11, 113-122.

Sosa, Ernest (1991): *Knowledge in Perspective*, Cambridge: Cambridge University Press.

Sosa, Ernest (1996): "Postscript to 'Proper Functionalism and Virtue Epistemology"', in: Jonathan L. Kvanvig (ed.), *Warrant in Contemporary Epistemology*, Totowa: Rowman & Littlefield, 269-278.

Sosa, Ernest (1999a): "How must Knowledge Be Modally Related to What Is Known?", in *Philosophical Topics*, 26, 373-384.

Sosa, Ernest (1999b): "How to Defeat Opposition to Moore", in *Epistemology*, 13, 141-153.

Sosa, Ernest (2000): "Skepticism and Contextualism", in *Philosophical Issues*, 10, 1-18.

Sosa, Ernest (2007): *A Virtue Epistemology*, *Apt Belief and Reflective Knowledge*, Oxford: Oxford University Press.

Sosa, Ernest (2009): *Reflective Knowledge*, *Apt Belief and Reflective Knowledge*, Oxford: Oxford University Press.

Sosa, Ernest (2015): *Judgment and Agency*, Oxford: Oxford University Press.

Stalnaker, Robert (1978): "A Defense of Conditional Excluded Middle", in: idem et al. (eds.), *Ifs*, *Conditionals, Belief, Decision, Chance and Time*, Dordrecht: Reidel, 87-104.

Stoutenburg, Gregory (forthcoming): "In Defense of an Epistemic Probability Account of Luck", in *Synthese*.

Stratton-Lake, Philip (2000): *Kant, Duty, and Moral Worth*, London: Routledge.

Sverdlik, Steven (2001): "Kant, Nonaccidentalness and the Availability of Moral Worth", in *Journal of Ethics*, 5, 293-313.

Swain, Marshall (1981): *Reasons and Knowledge*, Ithaca: Cornell University Press.

Thomson, Judith J. (1986): "Imposing Risks", in: idem (ed.), *Rights, Restitution, and Risk*, *Essays in Moral Theory*, Cambridge: Harvard University Press, 173-91.

Timmermann, Jens (2009): "Acting From Duty: Inclination, Reason and Moral Worth", in: idem (ed.): *Kant's Groundwork of the Metaphysics of Morals*, *A Critical Guide*, Cambridge: Cambridge University Press, 45-62.

Turri, John (2011): "Manifest Failure: The Gettier Problem Solved", in *Philosopher's Imprint*, 11, 1-11.

Unger, Peter (1967): "Experience and Factual Knowledge", in *Journal of Philosophy*, 64, 152-178.

Unger, Peter (1968): "An Analysis of Factual Knowledge", in *Journal of Philosophy*, 65, 157-170.

Unger, Peter (1970): "Our Knowledge of the Material World", in: Nicholas Rescher (ed.), *Studies in the Theory of Knowledge*, Oxford: Blackwell, 40-61.

van Riel, Raphael (2016): "Real Knowledge Undermining Luck", in *Logos & Episteme*, 7, 325-344.

Vogel, Jonathan (1987): "Tracking, Closure, and Inductive Knowledge", in: Steven Luper (ed.), *The Possibility of Knowledge, Nozick and His Critics*, Totowa: Rowman & Littliefeld, 197-215.

Wedgwood, Ralph (2016): "Objective and Subjective 'Ought"', in: Nate Charlow & Matthew Chrisman (eds.), *Deontic Modality*, Oxford: Oxford University Press, 143-168.

Whiting, Daniel (2017): "Against Second-Order Reasons", in *Noûs*, 51, 398-420.

Williams, Bernard (1981): "Moral Luck", in: idem (ed.), *Moral Luck*, Cambridge: Cambridge University Press, 20-39.

Williams, Bernard (1993): "Moral Luck: A Postscript", in: idem (ed.), *Making Sense of Humanity, And Other Philosophical Papers, 1982-1993*, Cambridge: Cambridge University Press, 241-247.

Williamson, Timothy (2000): *Knowledge and Its Limits*, Oxford: Oxford University Press.

Yang, Xiaomei (2000): "The Problem of Overdetermination: The Metaphysical and Epistemological Criteria of Moral Worth of Right Action", in *Southwest Philosophy Review*, 16, 73-86.

Zagzebski, Linda (1994): "The Inescapability of Gettier Problems", in *Philosophical Quarterly*, 44, 65-73.

Zagzebski, Linda (1996): *Virtues of the Mind, An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press.

Zhao, Haicheng (forthcoming): "Knowledge without Safety", in *Synthese*.

Zimmerman, Michael (2006): "Is Moral Obligation Objective or Subjective?", in *Utilitas*, 18, 329-361.