Dirty Hands and Moral Conflict – Lessons from the Philosophy of Evil

Christina Nick[1] ◍

Abstract
According to one understanding of the problem of dirty hands, every case of dirty hands is an instance of moral conflict, but not every instance of moral conflict is a case of dirty hands. So, what sets the two apart? The dirty hands literature has offered widely different answers to this question but there has been relatively little discussion about their relative merits as well as challenges. In this paper I evaluate these different accounts by making clear which understanding of concept distinctness underlies them and which of them is, ultimately, the most plausible one in the case of dirty hands and ordinary moral conflict. In order to do so, I will borrow from the terminology employed in recent debates in the philosophy of evil which have tackled a similar problem to the one at hand, i.e. defining what sets evil apart from ordinary wrongdoing. Here it has been argued that concepts could be distinct in three ways: they can have a quantitative difference, a strong qualitative or a moderate qualitative difference. I conclude that the most convincing definition of dirty hands draws a moderate qualitative distinction between ordinary moral conflict according to which dirty hands are those moral conflicts that involve a serious violation or betrayal of a core moral value.

Keywords Dirty hands · Evil · Moral conflict · Concept distinctness

1 Introduction

In a recent paper, Matthew H. Kramer has argued that "every problem of dirty hands is a moral conflict, but not every moral conflict is a problem of dirty hands" (2018, p. 187). The same issue had earlier been raised by Stephen de Wijze who pointed out that "not every wrong action is a dirty action and not every moral conflict involves dirty hands" (2007, p. 9). So, what sets the two concepts apart? In the following I

✉ Christina Nick
    C.Nick@leeds.ac.uk

1   Inter-Disciplinary Ethics Applied Centre, University of Leeds, Leeds, UK

will critically assess some of the criteria that have been proposed for distinguishing dirty hands from more ordinary types of moral conflict and will, in the end, suggest a novel approach to demarcate these two concepts.

Roughly speaking, an agent is faced with a moral conflict when they are confronted with two morally valuable options which can be performed separately but not jointly. As a result, the agent is forced to forego one of the moral values but does so for good moral reasons, i.e. to pursue the other morally valuable option.[1] The problem of dirty hands is concerned with a particular sub-set of such choices. The standard example given to illustrate dirty hands scenarios is Walzer's (1973) ticking bomb case. Imagine a politician who is committed to being morally good. She wants to act in the best interest of her citizens and adhere to her deep-rooted moral convictions. Immediately she is confronted with the following situation: the leader of one of the violent factions in the capital has been arrested and is thought to know the location of various bombs hidden in the city and set to explode by the end of the day. If the suspicions prove to be true, hundreds of innocent people will die in the explosions. Questioning has had no success, so our politician is now asked to authorise a more drastic measure. She decides to order the torture of the suspect in order to extract information about the location of the bombs. As a result, the politician gets her hands dirty; her choice generates a moral remainder and a particular negative emotional response that de Wijze has termed "tragic-remorse" (2004).[2] The crucial element in this emotion is a form of anguish and moral pollution on the side of the agent. Even though the politician chose the all-things-considered preferable option, she feels anguished about having to injure another human being seriously and that she chose to do so sticks with her; it pollutes her in a sense.

My aim of this paper is to understand how we can demarcate the concept of dirty hands from other kinds of moral conflict. Clarifying the concept of dirty hands has a variety of benefits. Firstly, we may hold that conceptual clarity in and of itself matters in analytic philosophy. Secondly, there has been some debate as to whether the problem of dirty hands is a rare occurrence or whether it is ubiquitous in our lives (Williams, 1978, pp. 61–62; Ramsay, 2000, pp. 38–39; Walzer, 2004, pp. 33–50; Tillyris, 2015, p. 64), and an improved understanding of the limits of the concept of dirty hands will help us provide an answer to this question. Finally, the problem of dirty hands is associated with a particular emotional response, so accurately demarcating dirty hands from ordinary moral conflict will help us to understand when it is appropriate to experience tragic-remorse. After all, tragic-remorse and a sense of pollution do not seem to be appropriate in all cases in which an agent has to choose between conflicting moral values. Imagine you are on the way to meet a friend for coffee. The day has been set for a long time, you have both been looking forward to seeing each other, and you promised that you would not miss it under any circumstances. On the way to the café you witness a traffic accident on a lonely country road. You rush over because both parties involved seem to have been seriously

---

[1] For an in-depth discussion of the nature of moral conflict see (Hurka, 1996).

[2] For further discussions on the sort of remainder and pollution appropriate to dirty hands see (Bagnoli, 2000; Baron, 1988; Williams, 1976).

injured. You call an ambulance and you decide to stay around to ensure that everyone is going to be okay. This, unfortunately, means that you will not make it to the café to meet your friend; you will have to break the promise that you gave to her. While we can see that this is a moral conflict in which you have to violate something of moral value but only do so to pursue something else of moral value, we would not want to say that you got your hands dirty by staying to help the victims of the accident rather than keeping your promise to your friend to go for coffee. It would not be appropriate for you to feel seriously anguished and morally polluted by your choice; instead, a form of regret may be more appropriate. So, what exactly is it that sets apart cases such as the ticking bomb scenario from cases such as breaking a promise to see a friend? In other words, what is the difference between the problem of dirty hands and ordinary moral conflict?[3]

The dirty hands literature has offered widely different answers to this question but there has been relatively little discussion about their relative merits as well as challenges (de Wijze, 2007; Kramer, 2018; Stocker, 1990; Walzer, 2004, 2006). In this paper I want to evaluate these different accounts by making clear which understanding of concept distinctness underlies them. In order to do so I will borrow from the terminology employed in recent debates in the philosophy of evil which have tackled a similar problem to the one at hand, i.e. defining what sets evil apart from ordinary wrongdoing (Calder, 2013; Kramer, 2014; Liberto & Harrington, 2016; Russell, 2014; Steiner, 2002). The word evil is commonly used in two different ways. The first, referred to as natural evil, construes the term in the broadest sense and refers to any kind of badness in the world. In this broad sense a natural disaster or even a toothache could be an evil. The second, referred to as moral evil, constructs the concept more narrowly so that only certain kinds of morally wrongful actions committed by responsible moral agents could be called evil. More specifically, evil in this sense refers to actions that in one way or another go above and beyond mere wrongdoing. For the purposes of this paper I will be concerned with the debates surrounding this narrow understanding of evil. The problem is that if evil is something above and beyond mere wrongdoing, then how exactly can we conceptualise this difference?

I will show in this paper that, following the recent debates in the literature on evil, concepts could be distinct in, at least, three different ways. Two concepts[4] can have a quantitative difference, a strong qualitative or a modest qualitative difference.[5] Under each of these headings I will consider and critically evaluate the various

---

[3] There are two ways in which we might want to clarify what is meant by a term. On the one hand we might wish to clarify what people refer to when they actually use the term, while on the other we might want to show what the term should correctly refer to and how people ought to use it. The aim of my paper is the latter.

[4] Whenever I talk about concepts being distinct in this way, it is merely a shorthand for saying that there is such a difference in the properties that the concepts represent.

[5] It is worth noting that recently a fourth way to distinguish two related concepts has been proposed by Liberto and Harrington (2016) that they call Quality of Emphasis or QE distinctness. No account of dirty hands appears to have drawn a distinction to ordinary moral conflict based on this approach, so I will not discuss this option here.

accounts of the difference between dirty hands and ordinary moral conflict that have been offered in the literature.[6] Ultimately, I will argue that the most convincing definition of dirty hands draws a modest qualitative distinction between ordinary moral conflicts according to which dirty hands are those moral conflicts that involve a serious violation or betrayal of a core moral value.

2 Quantitative Distinctness

Two related concepts can be quantitatively distinct when both concepts have the same properties, but one of the concepts has one or more of these properties to such an extent that it merits distinction. If, for example, evil was quantitatively distinct from wrongness, then if something is evil it is simply very, very wrong to such an extent that we are warranted in having a special term for it. Luke Russell, for example, embraces such an understanding when he argues that all "evil actions are extreme culpable wrongs". A culpable wrong is extreme if it "produces or was intended to produce […] an actual or possible harm that is extreme for at least one victim" (2014, p. 62).

   In a similar vein, we might want to say that dirty hands are quantitatively distinct from ordinary moral conflict in that dirty hands are serious violations of a value for good moral reasons. While ordinary moral conflict involves foregoing a value, dirty hands are those instances of moral conflict which involve a serious violation of that value. We could, for example, find such an account in a quantitative reading of Michael Stocker.[7] For Stocker the main distinguishing feature of dirty hands is that they "do not simply involve mere bads and harms. […] Rather, they involve betrayals and violations of people, principles or values" (1990, p. 25). In an ordinary moral conflict we might have to leave someone or something slightly worse off or wronged, but only in a dirty hands scenario would we go so far as to say that we have violated or betrayed someone or something. In this sense dirty hands and ordinary moral conflict would be quantitatively distinct because the latter involves serious as opposed to more minor infringements of a value. This interpretation of

---

[6] Dirty hands are often described as a choice between lesser evils, so trying to find answers to the definition of dirty hands in the literature on evil appears natural. We ought to be careful with our choice of words here, though, because there is not necessarily any conceptual link between dirty hands and evil. John Kekes for example, has argued that "exceptional circumstances demand that evil be done in order to prevent even greater evil" (2007, p. 207) specifically allowing for the possibility of evil actions that could, nonetheless, be the all-things-considered right choice. Others, however, have argued that any action that is a moral violation done for good moral reasons could never be called evil (Card, 2002, pp. 17–18; Singer, 2004, p. 186; Russell, 2014, p. 17). We could also imagine a position that is sceptical about the concept of evil but acknowledges the existence of dirty-handed actions. For the purposes of this paper I will therefore stay agnostic about the relationship between evil and dirty hands itself. For this reason I will not focus on Kramer's answer to the question I posed at the beginning of the paper because he defines dirty hands in virtue of evil actions.

[7] There is also another potential reading of his interpretation of the distinction between dirty hands and ordinary moral conflict. This reading has led Stephen de Wijze (2007) to construct a qualitative distinctness between dirty hands and ordinary moral conflict based on the circumstances in which the agent finds herself in. I will discuss this position in the next section.

Stocker's account is further supported by an earlier claim he makes. He states that "in paradigm cases of dirty hands, the costs are important moral costs" (e.g. killing someone to save one's country), but dirty hands can also "involve less important wrongs – e.g. to deceive the enemy into thinking that his loyalties have changed, a secret agent publicly insulting his old teacher" (1990, p. 18). These examples suggest that it is not so much the nature of the value that makes the difference, but rather the level of the violation or betrayal involved that sets dirty hands apart from ordinary moral conflict. Dirty hands are then distinct from other ordinary moral conflicts because they involve the serious violation of a moral value. Drawing a quantitative distinction can make sense of why the ticking bomb scenario is a paradigm example of dirty hands whereas the case of the traffic accident and the missed coffee date is not. When a politician orders the torture of a suspect this constitutes a serious violation of that suspect's bodily and mental integrity as well as their right to due process. On the other hand, while I have broken a promise to my friend to meet them this is hardly a serious violation of our friendship and their trust. It would feel odd, for example, to say that I have betrayed them by not showing up. So far, so good.

Within such a quantitative framework we are then left with two options about where to draw the dividing point between ordinary moral conflict and dirty hands. One possibility is that dirty hands constitute the extremity on a continuum of moral conflicts so that, a case of dirty hands is the most serious violation possible. There could therefore be no degrees of dirt; one action cannot be more or less dirty than another. The other option is that at some point on the continuum of moral conflicts there is a threshold beyond which actions are instances of dirty hands. There could therefore be more or less dirty actions depending on the extent to which they exceed this threshold. The latter option seems preferable because it can make better sense of the intuition that some actions can be dirtier than others.[8] Michael McDonald, for example, has argued that some actions are dirtier than others, so much so that we might even want to say that someone's hands are outright bloody rather than merely dirty (2000, p. 194). Additionally, similar to what Russell has argued in the case of evil, if we took dirty hands to constitute the very extremity on the moral conflict scale, then we would end up with a situation in which no action would be a case of dirty hands, because we could always imagine an action that would be slightly dirtier than the one in question (e.g. it would kill one more innocent human being). Therefore, on the quantitative view, dirty-handed actions are best conceived as those that lie beyond a certain threshold of severity on a continuum of moral conflict. This ensures that there are in fact cases of dirty hands and that we can distinguish between those that are more and less dirty.

However, a potential objection to arguing for a quantitative distinctness between any two concepts is that it will be difficult to determine exactly where this threshold lies. In the case of evil, Russell argues that we are faced with a vague concept and that we therefore ought to allow for a grey area in which intuitions conflict instead

---

[8] In the case of evil, Russell adopts this latter strategy as well and acknowledges that, "even though it is crass and disrespectful in some circumstances to rank evils, we should admit that some evil actions are morally worse than others" (2014, p. 65).

of attempting to draw a clear threshold. Similarly, in the case of dirty hands, it is unclear where exactly lies the threshold that makes foregoing a moral value for good moral reasons severe enough to count as a serious violation or a betrayal and therefore leaves us with a considerable grey area in which there will be disagreement on whether we are faced with a mere moral conflict or with dirty hands. Russell suggests that "perhaps the best we can do in these circumstances is […] to give examples of cases that clearly fall on either side of the dividing line" (2014, p. 67). I do not think that the conclusion that a concept is vague should worry us. Our everyday language is rife with vague concepts (e.g. red, tall) and it is not obvious that this is a problem in and of itself. Russell reminds us to acknowledge that "this kind of vagueness is to be expected in moral concepts, as in most other concepts" and that therefore "the fact that we cannot specify exactly the extremity […] linked to [a concept] does not indicate that there is any special problem with the concept" (2014, p. 67).[9] Additionally, in the case of dirty hands, we might think that such vagueness can actually make better sense of the way we talk about moral conflict in everyday life. We might call a certain action pure or clean when nothing of moral value has been foregone, but just because a decision is not entirely pure or clean does not imply that we would want to call it outright dirty. I think this is a plausible stance that we can take and one that certainly would fit with the phenomenology of our moral experience.

A more serious problem for quantitative approaches is that they appear to suggest that, ultimately, we could explain away one of the concepts to be distinguished. In the case of evil, Todd Calder has argued that "if evil is just very wrong, we can do without the term 'evil'. We can say everything we need to say using terms such as 'very wrong' or 'very very wrong'" (2013, p. 178). But if evil is just very wrong, Calder goes on, then the two concepts are not actually distinct from one another. This is an explanatory challenge according to which any quantitative account would struggle to provide us with a reason for why we have the term "evil" in the first place. In the case of dirty hands there would be a similar worry that drawing a quantitative distinction would imply that we could simply do away with the term "dirty hands", and that we could instead express the same thought by calling the case at hand a "serious moral conflict". A potential response would be that this is not something unique to the concepts of evil or dirty hands and that we have a variety of terms that we can substitute for a concept to indicate that it sits towards the extreme end of a continuum. For example, we can call something that is very, very small "miniscule" or "tiny" while we can call something that is very, very large "enormous" or "gigantic". Just because these concepts do not demarcate a qualitatively distinct concept does not mean that we would automatically have to do away with them. Furthermore, whatever explanation we have for continuing to use these terms

---

[9] And even if it would turn out to be a problem, this problem would be equally shared by any qualitative approach (i.e. an approach that thinks that a concept is distinct in virtue of possessing a unique property not had by the other) because these will themselves be relying on properties that have a vague threshold somewhere on a continuum. For example, if we argue that evil is to commit a wrong out of sadistic malice the qualitative approach will still have to show at what point a motivation has reached the required threshold to be called sadistic malice.

in everyday language should then also apply to the case of using the terms "evil" and "very, very wrong" or "dirty hands" and "serious moral conflict". There is therefore nothing unusual or concerning about having a variety of terms that can be substituted for one another.

This response, however, does not quite go to the bottom of the explanatory challenge. The real motivation behind that objection is to bring out our intuition that to call something "very, very wrong" does not convey the same sort of meaning as calling something "evil", i.e. that they cannot be straightforwardly substituted after all. To say, for example, that what Hitler did was very, very wrong[10] would seem to fall short of an adequate description of the horrors of the Holocaust. To say that what Hitler did was evil appears to be a more suitable locution. But, how, if evil and wrongdoing are merely quantitatively distinct, could we explain this? I think we might have similar intuitions regarding cases of dirty hands involving killing, torturing or otherwise violating others' basic human rights. It would not be enough to say that what the agent had to do was to, regrettably, forego a moral value, instead we want to say that her actions were "dirty". The notion of "dirt" brings with it a connotation of a remainder and pollution that cannot be captured in the same way by saying that something was a "serious" moral violation. But how, if dirty hands and ordinary moral conflict are merely quantitatively distinct could we explain this?

In response to this a defender of the quantitative view could argue that while two quantitatively distinct terms are semantically equivalent (e.g. "evil" and "very, very wrong" are both used to describe an action as lying beyond the specified threshold on the wrongness scale), one of them can have, what Paul Grice has termed, a "conventional implicature" (1989, pp. 21–57) that is missing in the case of the other (e.g. when we say that something is "evil" we imply that it is one of the worst things imaginable).[11] So in response to the problem that saying "what Hitler did was very, very wrong" appears to convey a different meaning to saying "what Hitler did was evil", the defender of the quantitative approach could argue that, while "evil" and "very, very wrong" have the same meaning, they have different implicatures. This line would appear to be compatible with Russell's approach, who argues that the term "evil" has an "unparalleled ability to capture and to convey our wholehearted condemnation of an action" (2014, p. 37). The core features, according to Grice, of a conventional implicature are that a) it cannot be calculated and b) it is detachable but not cancellable. As opposed to a conversational implicature, which can be understood by considering what has been said in the particular context of the utterance, a conventional implicature cannot be made sense of without knowledge of the particular conventional meaning attached to the term in question. Secondly, conventional implicatures are detachable because, if the quantitative difference view is right, we can simply exchange

---

[10] Other intensifiers are faced with the same problem. For example, to say that what Hitler did was extremely wrong would still not seem to capture the same meaning as saying that what Hitler did was evil.

[11] Thank you to Will Gamester for suggesting this as a potential response that a quantitative account could offer.

the term "very, very wrong" for the term "evil" without altering the truth conditions of the sentence. At the same time, however, conventional implicatures are not cancellable because once the term is uttered we are unable to distance ourselves from what it implicates; e.g. it would not make sense to say that "what Hitler did was evil, but I don't want to imply that his actions were some of the worst things imaginable". Similarly, one might wish to argue that for those engaged in the debate, the term "dirty hands" has a conventional implicature missing from the term "moral conflict". This is to say that when someone says that a case is an instance of dirty hands this brings with it the implicature that it is the sort of case that will result in anguish and a sense of pollution for the agent. While interchanging the terms "dirty hands" and "serious moral conflict" would not change the truth conditions of the sentence, once we use the term "dirty hands" we cannot but implicate that the agent ought to be anguished and will be polluted by having committed a moral violation for good moral reasons.

The problem with relying on the notion of a conventional implicature to respond to Calder's explanatory challenge is that the idea of a conventional implicature itself is controversial and leaves us with as many questions as it provides us answers. The main challenge that a proponent of this approach faces is to explain how something can be implicated by our conventional use of a term without this implication being part of the term's meaning. The problem of doing so can be shown when we put purported conventional implicatures through the indirect quotation (IQ) test proposed by Bach (1999, p. 340). Imagine that your friend Marv states that "Shaq is huge but he is agile". According to a conventional implicature analysis, the truth-conditions of Marv's utterance are met when it is true that Shaq is both huge and that he is agile. While the two sentences are semantically equivalent, "but" appears to have the additional implicature that there is some form of contrast between being huge and agile such that it is maybe unusual or surprising for someone to display both characteristics. We can see the problem with this analysis when we consider how someone may indirectly quote Marv. If Sally reports that "Marv said that Shaq is huge but that he is agile" she accurately reflects what Marv has said. Should she, however, report that "Marv said that Shaq is huge and that he is agile", her representation of Marv's utterance is incomplete. The contrasting content of "but" therefore cannot be detachable but instead is part of what is being said (1999, p. 399). Another illustration of this problem can be found in Christopher Hom's paper on pejoratives in which he criticises accounts that make sense of slurs in terms of conventional implicatures. If speaker A says that "Institutions that treat Chinese as chinks are racist" and B reports that "A said that institutions that treat Chinese as Chinese are racist", B does not accurately reflect what A said. The pejorative content of the slur therefore cannot be detachable but instead is part of what is being said (Hom, 2008, p. 425).

While I cannot engage with the literature on conventional implicatures in more depth here, given that some of the most common purported examples of conventional implicatures fail, I think this puts the onus on defenders of a quantitative approach to explain how "evil" and "very, very wrong" or "dirty hands" and "serious moral conflict" can be semantically equivalent while "evil" or "dirty hands" contain an additional conventional implicature that is somehow detachable and not part of the term's semantics.

3 Qualitative Distinctness

There are two different approaches to drawing a qualitative distinction between dirty hands and ordinary moral conflicts that have been taken in the literature. The first one adds what I will call an "external immorality condition" while the second adds a "core value condition" to distinguish dirty hands from ordinary moral conflict. I will take each of these approaches in turn.

3.1 External Immoralities

What does it mean for two related concepts to be qualitatively distinct from each other? In the context of evil, Hillel Steiner argues that it is "the addition of some property that, in association with the wrong-making properties of the thing involved, renders it evil" (2002, p. 184). One interpretation of this claim would be that two related concepts are qualitatively distinct if and only if one of them possesses a property not possessed by the other to any degree. To illustrate this, he provides the example of aggravated assault which is qualitatively distinct from ordinary assault because it possesses the additional quality of an intent to murder. Steiner argues that the particular property that distinguishes evil from ordinary wrongdoing is that in the former the agent takes pleasure in the wrongdoing (2002, p. 189). I will follow Russell in calling two concepts that can be distinguished in this way "strongly qualitatively distinct" (2014, p. 116).

In a similar vein, we could argue that dirty hands possess a property not shared by ordinary moral conflicts. We could argue, for example, that dirty hands involve an agent being forced by the immoral actions of others to forgo a value for good moral reasons. According to this view, whenever an agent is faced by a moral conflict that was created by the immoral acts of another, we call such cases a problem of dirty hands and not merely an ordinary moral conflict. In this case we have added an external immorality condition that distinguishes dirty hands from other kinds of moral conflict because the former possess a property entirely absent in the latter. de Wijze (2007) advocates for such an external immorality condition based on Stocker's (1990) analysis of dirty hands.[12] Following from the analysis I laid out in the previous section, Stocker continues to argue that "the immorality of the circumstances *can* provide the specific difference between cases of dirty hands and other cases" (1990, p. 19).[13] More specifically, in "many of the clearest cases of dirty hands […] the agent is immorally coerced to take part in, perhaps even to help implement, an immoral project" (1990, p. 20). It is important to note here that Stocker himself does not think that this is a necessary criterion and that it is possible to imagine cases of dirty hands that do not involve "immoral coercion" (1990, p. 25). Though he does acknowledge that this will often be difficult, especially when thinking about the most serious kinds of dirty hands conflicts. De Wijze then goes the additional

---

[12] As I will point out later, de Wijze's account is more complex than this. For the moment, however, I will only concentrate on this particular element of his definition.

[13] My emphasis.

step and argues that *all* cases of dirty hands are "due to the immoral circumstances created by other persons (or organisations of persons) within which an agent finds herself" (2007, p. 12).

Drawing a strong qualitative distinction between two concepts will, however, often be problematic. In the case of evil, Russell argues against using such a distinction because "while […] it would be advantageous to draw a sharp line between evils and non-evil wrongs, ceteris paribus, […] we should not endorse the strong version of the qualitative difference thesis, because doing so would commit us to implausible claims about the extension of the concept of evil action" (2014, p. 119). The problem with Steiner's account, for example, is that it would include trivial wrongdoings such as shoplifting in the category of evil as long as the agent was taking pleasure in them. This seems to contradict how we would commonly use the notion of evil. The general problem for accounts that want to draw a strong qualitative distinction between related concepts is that they will have difficulty trying to find a single property that is always entirely absent from one of them. A similar problem is faced by anyone wanting to use the external immoralities condition to draw a strong qualitative distinction between dirty hands and ordinary moral conflict. Imagine that I have a horrible boss. She decides to give me too much work because she derives pleasure from seeing me anxious and struggling to achieve a healthy work-life balance. As a result, I get home late, where I am faced with the following choice: either I can read my child a bedtime story or I can help my partner with some much-needed housework. The immoral actions of my boss have pushed me into a moral conflict in which I have to choose between the value of caring for and bonding with my child and the value of ensuring that the burden of the unpaid labour of household tasks is divided equally between me and my partner. While this certainly is a difficult situation, I don't think that we would want to say that I have dirty hands for choosing to read a bedtime story rather than doing the washing up. As such, the external immorality condition does not seem to be unique to dirty hands and therefore cannot be used to draw a strong qualitative distinction between dirty hands and ordinary moral conflicts.

To avoid this problem, Russell proposes that a strong qualitative account could include an additional "extremity condition" (2014, p. 119). In the case of Steiner's account of evil (in which evil is distinguished from ordinary wrongdoing in that it possesses the property of "taking pleasure in the action") this would ensure that only extreme wrongdoings that are sufficiently harmful and in which the agent takes pleasure can cross the threshold required for an action to be evil. The problem, Russell argues, is that as soon as we add such a condition, Steiner's account can no longer claim that there is a "qualitatively distinct psychological property possessed by all and only evil actions" (2007, p. 670). After all, a shoplifter who takes pleasure in stealing a bar of chocolate would no longer be committing an evil action on this revised account. We therefore "ought to admit that sometimes the difference between an evil action and a non-evil action is simply that the evil action inflicted much more pain" (2014, p. 121).[14]

---

[14] Matthew Kramer has recently argued that we should abandon the quantitative/qualitative dichotomy and instead embrace that all concept distinctness will ultimately rely on both quantitative as well as qualitative difference. Evil for him is conduct connected to severe harm and motivated by an evil mentality.

Todd Calder, however, has argued that Russell is mistaken in thinking that adding an extremity condition implies that the distinction between concepts is no longer exclusively qualitative. To show this he invites us to think about the conceptual distinction between tables and chairs. Intuitively we think that there is a qualitative distinction between tables and chairs but, were we to accept Russell's interpretation that two concepts can only be genuinely qualitatively distinct if one of them has a unique property not possessed by the other to any degree, we would have to conclude that tables and chairs are not in fact qualitatively distinct. Both tables and chairs can have four legs, have hard flat surfaces, be backless, be made from durable materials etc. Therefore, neither tables nor chairs possess any unique property that can never be had by the other and, according to Russell's argument, could not be qualitatively distinct. Calder argues against this by pointing out that while tables and chairs might not have any unique property not possessed by the other, what distinguishes them is that they have differing essential properties. According to Calder, "the essential distinguishing property of a chair is that it is primarily used for sitting on, while the essential distinguishing property of a table is that it is primarily used for setting things on" (2013, p. 180). Neither of these essential properties is exclusive to either tables or chairs; we can, after all, sit on a table and put things onto a chair. Calder therefore concludes that a concept can be "qualitatively distinct from another provided that all of the essential properties of the first concept are not all of the essential properties of the second concept, but had to a greater or lesser degree" (2013, p. 181). Calder then proposes that evil has at least two essential properties: one is that it causes significant harm understood "as a harm that a normal rational human being would take considerable pains to avoid", while the other is that the agent has an "e-motivation" understood as "an inexcusable intention to bring about, allow, or witness the significant harm […] for an unworthy goal" (2013, p. 188). According to Calder, ordinary wrongdoing does not have either of these properties as an essential feature, and therefore evil and wrongdoing are qualitatively distinct. After all, it is not essential for wrongdoing that I cause significant harm (e.g. I could be a lucky drunk driver who does not actually cause an accident), nor is it necessary for wrongdoing that I have an inexcusable intention (e.g. I could be an absent-minded construction worker who does not sufficiently secure a dangerous building site thereby failing to prevent a fatal accident). To distinguish this understanding

---

Footnote 14 (continued)

Because of the requirement for an evil mentality, he argues, the concepts of evil and mere wrongdoing are genuinely qualitatively distinct. It is an essential property not held by ordinary wrongdoing. But, these different mentalities "are themselves vague properties […] that is, each of them is differentiated from a more benign type of outlook through a series of fine gradations, and no gradation is a talismanic point of transition between the benign type and the potential evil type" (2018, p. 193). Ultimately, the concepts of evil and ordinary wrongdoing are therefore based on "a qualitative divide that emerges from quantitative gradations" (2014, p. 61). While I cannot attempt to give a thorough rebuttal of this position here, I think that Kramer unhelpfully muddies the waters in this debate. Even if vague concepts can be qualitatively distinct as a result of quantitative gradations, this does not get to the heart of the debate on how to define evil, i.e. whether evil is simply a form of severe wrongdoing or whether evil is a form of (severe) wrongdoing coupled with another property.

of qualitative distinctness from Steiner and Russel's strong version, Calder refers to this theory as "modest qualitative distinctness" (2019).

Going back to the external immorality condition, de Wijze's account is, in fact, more sophisticated than an attempt to draw a strong qualitative distinction. Next to the external immorality condition he also thinks that all dirty hands cases involve the extremity condition argued for by Stocker. For de Wijze, what all cases of dirty hands have in common is that they "involve the justified betrayal of persons, values or principles due to the immoral circumstances created by other persons (or organisations of persons) within which an agent finds herself" (2007, p. 12). The first part, following Stocker, suggests that only serious violations or betrayals of values[15] can count as instances of dirty hands while the second part ensures that only those instances that have been created by the immoral actions of a third party can be called dirty hands cases. On this account, dirty hands and ordinary moral conflict would be modestly qualitatively distinct. While neither the extremity nor the external immorality condition are unique properties of dirty hands, only for dirty hands are they essential. In order to show that they are not unique to dirty hands, all we need to do is to consider cases of ordinary moral conflict that possess either of these properties. We can see that the extremity condition is not unique to dirty hands if we consider a case of conflict in which we seriously violate a moral value but in which we do not think that we get our hands dirty. Imagine, for example, the captain of a ship who has an obligation (both legal and moral) to their client to safely transport their goods across the sea. Now the ship finds itself in the middle of a storm and the only way to save the crew is to cut loose the cargo and let it fall into the sea. While this a serious violation of their obligation, I do not think that we would want to say the captain now has dirty hands for saving their crew by abandoning the goods. We have already seen in the example of the horrible boss that the external immorality condition is not unique to dirty hands cases, either. To show that these two properties are only essential to dirty hands, what we need to do is consider cases of ordinary moral conflict that do not involve either the extremity or the core value condition. As we have seen in the beginning of the paper when considering the case of not meeting your friend for coffee in order to stay with the victims of an accident, a serious violation or betrayal of a value is not essential to ordinary moral conflicts. Similarly, it also does not seem essential to ordinary moral conflicts that they be caused by an external immorality. Imagine, for example, that the car accident that is causing me to miss my coffee date was not caused by reckless driving but by one of the drivers having a heart attack. While my moral conflict remains, it was not caused by the immoral

---

[15] There is a legitimate worry here that any modest qualitative account of dirty hands that uses the idea of a "violation" or "betrayal" to help demarcate the concept from ordinary moral conflict will appear rather ad hoc – that is, it will simply judge something to be a violation or betrayal when we already have the intuition that the case is an instance of dirty hands—unless we give an account of what it means for something to count as a "violation" or "betrayal". Stocker does not provide any clarification on this point. For the purposes of this paper in which I merely want to assess the initial plausibility of a variety of accounts it will be sufficient to say that the concepts are vague and that there will be a considerable grey area. A more sustained defence of any particular view involving the extremity condition will, however, need to say more about what constitutes a "violation" or "betrayal".

actions of a third party. Dirty hands and ordinary moral conflict would therefore be modestly qualitatively distinct.

While this account could, in principle, enable us to draw out two genuinely distinct concepts, I do not think that it can actually make sense of all cases of dirty hands. Imagine that you own two large dogs who have been your companions for several years and to whom you feel a strong obligation. Unfortunately, one night, your house catches on fire. Both of the dogs are scared and confused and are refusing to follow you out of the house. Because of their size you can only carry one of them at a time. Once you have rescued one of them, however, the fire will have spread so much that you will not be able to go back and rescue the second.[16] Would the cause of the fire, for example, whether it is caused accidentally by lightning or by an arsonist who laid the fire with the intention of seeing the anguish this choice would cause you, change the kind of remainder you feel and therefore indicate that one of them is a case of dirty hands while the other is not? I do not see a reason why you should feel the tragic-remorse that is indicative of a dirty hands problem if the fire was caused by an arsonist, but only something such as, for example, agent-regret[17] which is indicative that the case is not a dirty hands problem when the fire was caused by lightning. As de Wijze himself notes, agent- regret is a "form of regret which factors into our emotional responses our role in bringing about certain states of affairs albeit unintentionally" (2004, p. 462). He goes on to argue that, while agent-regret is an appropriate emotional response to cases in which we, without planning or wanting to, become causally involved in wrongdoing, it is not appropriate for cases in which "it is not merely the fact of [our] causal role in the event that is problematic, but [our] willing endorsement of a moral violation" (2004, p. 462). In such cases a form of remorse will be more appropriate because it instead focusses on "the agent's voluntary role in performing, or participating in, or allowing a particular action that results in a moral violation" (2004, p. 460). Given that, in both versions of our case in question, the agent consciously deliberates and chooses to prioritise one life over another, I do not see how the fact that one decision was caused by lightning while the other was caused by an arsonist could provide us with reasons to think that tragic-remorse would be appropriate in one case but only agent-regret in the other. As a result, I do not think that we have good reasons to believe that one is a case of dirty hands while the other is an instance of a more ordinary moral conflict. While I do not want to argue that there can never be a connection between the weightiness of the remainder and the circumstances that lead to the agent being faced with a moral conflict (e.g. we might feel that our hands are dirtier than they would have been had the situation been caused by an accident or natural causes), immoral circumstances

---

[16] Someone might argue that this kind of scenario looks more like an instance of a tragic dilemma in which there does not appear to be a lesser evil for us to choose, rather than a dirty hands problem in which one option presents itself as being slightly less objectionable than the other. Such a distinction appears to have been suggested by Bernard Williams (1978, pp. 61–62) but has been rejected by de Wijze (2007, pp. 8–9). For the purposes of this paper I will side with de Wijze in this debate and assume that dirty hands cases can involve both moral conflicts as well as tragic dilemmas.

[17] In response to an earlier draft of this paper de Wijze has suggested that this would be the more appropriate emotional response in the lightning case.

are simply not necessary to make something a case of dirty hands. For this reason, I think we should reject de Wijze's external immorality condition.

3.2 Core Value Condition

As a result, I want to consider an alternative approach for drawing a qualitative distinction between dirty hands and ordinary moral conflict. On this approach we could argue that the concepts are strongly qualitatively distinct because dirty hands involve foregoing a core moral value for good moral reasons. The argument goes that whenever such core moral values are involved, we would not speak of an ordinary moral conflict but always of a problem of dirty hands. What are core moral values? Take, for example, Michael Walzer's later writings on the subject of dirty hands and supreme emergencies. He conceptualises dirty hands situations as those in which a "rights normality" is overridden by a "utilitarianism of extremity" (2004, p. 40). Politicians usually ought to act in accordance with standard deontological constraints, but in cases of supreme emergency they should be willing to violate deontological constraints and act based on a utilitarian calculation instead. According to Walzer "a supreme emergency exists when our deepest values and our collective survival are in imminent danger" (2004, p. 35). Dirty hands are those instances in which we have to forgo an important deontological principle to secure the "ongoingness" of a society and its "way of life" (2004, p. 43). For Walzer, "the survival and freedom of political communities […] are the highest values of international society" (2006, p. 253). Developing a core value condition based on Walzer's writings on supreme emergencies is highly restrictive, though, and does not align well with some of the archetypical examples given to illustrate when an agent gets her hands dirty. For example, the ticking-bomb scenario introduced by Walzer himself in his earlier work would in most instances no longer count as a case of dirty hands, unless nothing short of the continued survival of an entire population were at stake. A successful account along this approach should therefore be more inclusive in the way it construes the notion of core moral values.

I think that a better strong qualitative distinctness account using a core value condition would argue that the difference between dirty hands and ordinary moral conflict lies in the strength of the commitment that an agent ought to have to the values that she is violating in a given moral conflict. For instance, a promise to meet up with a friend for coffee is something that we should value, but not something that we should be committed to so deeply that, in case we have to break it, it ought to anguish us and make us feel as if we had abandoned one of our most fundamental moral commitments. Torturing another human being, however, ought to have exactly this effect on us. What sets dirty hands apart from other forms of moral conflicts is that they force us to abandon some of what ought to be[18] our most deeply held moral

---

[18] I'm saying "ought to be" to rule out any odd, deeply held commitments people might have. For instance, imagine someone who thinks that the most important value in life is the pleasure of eating chocolate ice cream and who would experience serious anguish if they had to forgo an opportunity to eat more chocolate ice cream in order to save someone's life. We would not want to say that by foregoing the pleasure of eating chocolate ice cream in order to save someone's life they had to get their hands dirty.

commitments. The kind of situations I have in mind here are conflicts in which we are faced with having to, in some way, restrict or undermine the basic human rights of a person or the basic rights of non-human animals or destroy the environment we all need in order to flourish. It would be beyond this paper for me to lay out and defend an exact list of what these core values are. For the time being it suffices to note that while we may disagree about the exact makeup of this list, this does not undermine the fact that the core value condition can tell us a story about dirty hands that makes better sense of cases such as, for instance, the earlier dog rescue examples than the external immorality condition.

In order to show that there is a strong qualitative distinction between dirty hands and ordinary moral conflict we would have to show that there are no ordinary moral conflicts that involve a violation of one of these core moral values. Let us go back to my earlier example of the horrible boss. We can now imagine that my boss has developed a new tactic; she now wants to make me unpopular with my colleagues. She knows that my current work environment is having a negative impact on my well-being and that, as a result, I am desperate to find a new job. She tells me that, in order to get the favourable end-of-year review that I require for any successful application, I will have to go around the office and repeatedly pinch my colleagues' arms. Each pinch will cause a brief moment of mild pain, but especially due to its repeated nature, my colleagues will be inconvenienced and annoyed at the wantonness of my action (my boss has obviously forbidden me to tell my colleague the reason for pinching them). Respecting other people's bodily integrity and refraining from arbitrarily subjecting them to pain definitely ought to be among our core moral commitments, but in this instance the infringement seems so trivial that it would be odd to say that I have dirty hands for having pinched someone's arm. The core value condition therefore does not seem to be unique to dirty hands and therefore cannot be used to draw a strong qualitative distinction between dirty hands and ordinary moral conflicts.

Like the move made in de Wijze's account previously, we could add an extremity condition to this account in order to overcome this problem. We could draw out an account along these lines from a particular reading of Stocker. He is clear that values have to be violated or betrayed rather than an agent merely leaving someone slightly worse off or wronged in order for an action to be a case of dirty hands. Additionally, however, he also suggests in some places that what is violated or betrayed is a person or a particularly important value. He suggests that dirty hands conflicts arise when we face situations in which people will be "wronged, they and their trust, integrity, and status as ends […] violated, dishonoured, and betrayed" (1990, p. 17). Additionally they can also involve the destruction of items of great worth (e.g. art or a holy place) or "the violation of an important principle rather than a person" (1990, p. 18). (e.g. circumventing justice). While Stocker himself does not explicitly say so, we could combine these features so that dirty hands conflicts arise when we seriously violate or betray a person, important value or principle. This is in essence a combination of an extremity and a core value condition. Dirty hands and ordinary moral conflict, on this account, would be modestly qualitatively distinct because while neither of these two properties are unique to dirty hands, for dirty hands they are jointly necessary conditions. We have already seen, when considering the

external immorality condition, that there can be cases of ordinary moral conflict that fulfil the extremity condition without being cases of dirty hands. This was the case of the captain who jettisons goods in order to safe their crew. We have also already seen in the second example of the horrible boss that the core value condition is not unique to dirty hands cases. To show that these two properties are only essential to dirty hands, what we need to do is consider cases of ordinary moral conflict which do not involve either the extremity or the core value condition. As we have seen in the beginning of the paper when considering the case of not meeting your friend for coffee in order to stay with the victims of an accident, a serious violation or betrayal of a value is not essential to ordinary moral conflicts. Similarly, it also does not seem essential to ordinary moral conflicts that they violate a core moral value. Imagine, for example, that someone is forced to decide between helping an elderly person carry home their shopping and saving a child from drowning. By saving the drowning child instead of assisting the elderly person they have not violated a core moral value. Dirty hands and ordinary conflict would therefore be modestly qualitatively distinct. I think that this reading of Stocker is the most promising definition of dirty hands.[19]

4 Conclusion

I have shown that, while drawing a quantitative distinction between dirty hands and ordinary moral conflict is possible, it would ultimately mean that we could do away with the notion of dirty hands. In order to avoid this conclusion, we could attempt to show that dirty hands are strongly qualitatively distinct from ordinary moral conflict in that the former possesses a property that is entirely absent in instances of the latter. We considered that what is unique to dirty hands could be that they are caused by the immoral actions of others or that they involve foregoing a core moral value. The problem is that both options allowed for more trivial moral infringements, such as not helping with housework or pinching someone's arm, to count as instances of dirty hands. To avoid this, it would therefore make sense to add an extremity condition ensuring that more trivial instances of moral conflict could not count as getting one's hands dirty. On this approach, even if instances of ordinary moral conflict can share the properties of dirty hands, as long as the essential properties of the latter are distinct from the essential properties of the former, dirty hands can be modestly qualitatively distinct from ordinary moral conflict. We showed that the version arguing that dirty hands are brought about by the immoralities of a third party that force us to seriously violate a value draws questionable distinctions between cases that intuitively ought to be treated

---

[19] It might be worthwhile pointing out that there appears to be some common ground between the core morality condition I defend and the external immoralities condition defended by de Wijze. After all, the acts of others tend to be malicious because they violate some of our most important moral commitments. The core morality condition differs, however, in that it takes that violation to be what is fundamentally the dirt in dirty hands, whereas the external immoralities condition would suggest that it is complicity in another's malicious actions that constitutes this dirt.

alike. As a result, I suggested that the most promising way of defining dirty hands is that they are grave moral conflicts that involve the serious violation of a core moral value.

Authors' Contributions Not Applicable

Data Availability Not Applicable

Code Availability Not Applicable

Declarations

Conflicts of Interest/Competing Interests Not Applicable

References

Bach, K. (1999). The Myth of Conventional Implicature. *Linguistics and Philosophy, 22*(4), 327–366.

Bagnoli, C. (2000). Value in the Guise of Regret. *Philosophical Explorations, 3*(2), 169–187.

Baron, M. (1988). Remorse and Agent-Regret. *Midwest Studies in Philosophy, 13*(1), 259–281.

Calder, T. (2013). Is evil just very wrong? *Philosophical Studies, 163*(1), 177–196.

Calder, T. (2019). Evil and wrongdoing. In T. Nys & S. de Wijze (Eds.), *The Routledge Handbook of the Philosophy of Evil.* (pp. 218–233). Routledge.

Card, C. (2002). *The Atrocity Paradigm: A Theory of Evil.* Oxford University Press.

de Wijze, S. (2004). Tragic-Remorse - The Anguish of Dirty Hands. *Ethical Theory and Moral Practice, 7*(5), 453–471.

de Wijze, S. (2007). Dirty Hands: Doing Wrong to do Right. In I. Primoratz (Ed.), *Politics and Morality.* (pp. 3–19). Palgrave Macmillan.

Grice, P. (1989). *Studies in the Way of Words.* Harvard University Press.

Hom, C. (2008). The Semantics of Racial Epithets. *Journal of Philosophy, 105*(8), 416–440.

Hurka, T. (1996). Monism, Pluralism, and Rational Regret. *Ethics, 106*(3), 555–575.

Kekes, J. (2007). *The Roots of Evil.* Cornell University Press.

Kramer, M. H. (2014). The Nature of Evil. *The American Journal of Jurisprudence, 59*(1), 49–84.

Kramer, M. H. (2018). Problems of Dirty Hands As a Species of Moral Conflicts. *The Monist, 101*(2), 187–198.

Liberto, H., & Harrington, F. (2016). Evil, wrongdoing, and concept distinctness. *Philosophical Studies, 173*(6), 1591–1602.

McDonald, M. (2000). Hands: Clean and Tied or Bloody and Dirty? In P. Rynard & D. P. Shugarman (Eds.), *Cruelty and Deception.* (pp. 187–198). Broad View Press.

Ramsay, M. (2000). Democratic Dirty Hands. In L. Cliffe, M. Ramsay, & D. Bartlett (Eds.), *The Politics of Lying - Implications for Democracy*. (pp. 27–42). Macmillan.

Russell, L. (2007). Is Evil Action Qualitatively Distinct from Ordinary Wrongdoing. *Australasion Journal of Philosophy, 85*(4), 659–677.

Russell, L. (2014). *Evil: A Philosophical Investigation*. Oxford University Press.

Singer, M. (2004). The Concept of Evil. *Philosophy, 79*(308), 185–214.

Steiner, H. (2002). Calibrating Evil. *The Monist, 85*(2), 183–193.

Stocker, M. (1990). *Plural and Conflicting Values*. Oxford University Press.

Tillyris, D. (2015). "Learning How Not to Be Good": Machiavelli and the Standard Dirty Hands Thesis. *Ethical Theory and Moral Practice, 18*(1), 61–74.

Walzer, M. (1973). Political Action: The Problem of Dirty Hands. *Philosophy & Public Affairs, 2*(2), 160–180.

Walzer, M. (2004). *Arguing About War*. Yale University Press.

Walzer, M. (2006). *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. Basic Books.

Williams, B. (1976). Moral Luck. *Proceeding of the Aristotelian Society, Supplementary Volumes, 50*(1), 115–135.

Williams, B. (1978). Politics and Moral Character. In S. Hampshire (Ed.), *Public and Private Morality*. (pp. 55–73). Cambridge University Press.