

Uppsala University  
Department of Philosophy

# **Moral Worth, Moral Awareness, and Virtuous Motives**

Fabian Lorentzon

Master's Thesis in Practical Philosophy

Autumn Term - 2021

Supervisor: Olle Risberg

## 1. Introduction

Morally desirable actions do not always reflect well on those who perform them. It is morally desirable that a shopkeeper is honest to his customers. But he is not morally praiseworthy if he is only honest because it is profitable and would lie if that was instead more profitable (Kant 1785 [2018]). When evaluating whether someone's actions merit praise, the agent's motives appear central. What should the morally exemplary agent care about? Perhaps the shopkeeper should have cared about honesty for its own sake, or about his customers' well-being or about doing what is morally right. Whatever the answer is, it will indicate what actions merit the most amount of praise. Those actions are what philosophers have called *morally worthy*.

In the literature about moral worth, two competing ideas are usually discussed. The first is that what is important for an action's having moral worth is that the agent is motivated by a moral concern, such as a concern for doing what is morally right. That is, the agent's object of concern must contain a moral concept. Call this the *Moral Concept View*. The second is that what is important for an action's having moral worth is that the agent is motivated by concern for the features that *make* the action morally right, such as a concern for promoting someone's well-being. Here, the agent's object of concern will not contain a moral concept. Call this the *Non-Moral Concept View*.

In this paper I undertake three tasks. The first is to argue that the prospects for the Moral Concept View are highly limited. I do so by presenting a series of counterexamples. The second is to examine an important objection to the Non-Moral Concept View. The objection is that the Non-Moral Concept View is incompatible with the fact that morally worthy actions cannot be instances of people accidentally acting rightly. Unfortunately, previous responses to this objection have been lackluster. My diagnosis is that this is because of a misunderstanding of what it takes to act rightly non-accidentally. I will clarify this matter and remedy previous mistakes. The third is to draw on the previous sections to present a novel account of moral worth. As the discussion will indicate, morally worthy action not only involves a motivational component, but also a belief-component. I argue that the motivational component is accounted for by the Non-Moral Concept View, while the belief-component is not. Although the prospects for the Moral Concept View are highly limited, at least *implicit* moral beliefs remain important for moral worth.

The paper is structured as follows. After making some clarificatory remarks in section 2, I go on to criticize the Moral Concept View in sections 3 and 4. In section 5, I

present the objection that the Non-Moral Concept View is incompatible with the fact that morally worthy actions cannot be instances of people accidentally acting rightly and critically examine some attempted responses. In section 6, I diagnose the underlying reason why previous responses have failed and suggest what the proper way forward is. In section 7, I draw on the previous sections to motivate my preferred account of moral worth. On this account, an action is morally worthy just in case (1) it is morally right, (2) it is performed from concern for the relevant right-making features, and (3) the agent's motivation to perform the action is explained by his recognition of the relevant moral facts. In section 8, I respond to questions and objections to my account. The paper concludes in section 9 with some final thoughts.

## 2. Clarificatory remarks

At first, I will make some clarificatory remarks. This paper will be concerned with morally praiseworthy actions, and more specifically, the question of morally worthy actions. The question of moral worth closely relates to the question of praiseworthiness, and it is difficult to establish the exact difference between an action being morally worthy and it being morally praiseworthy. However, there is some consensus that saying that an action is morally worthy is stronger than saying that it is praiseworthy. For example, Julia Markovits (2010, p. 237-241) argues that morally worthy actions are the *most* praiseworthy actions. Similarly, Alison Hills states: "Morally worthy action is the *best* kind of action, the *most* praiseworthy and morally admirable" (emphasis added) (Hills 2018, p. 261). In this paper, I adopt this interpretation of moral worth. I will assume that praiseworthiness comes in degrees and that to say that an action has moral worth means that the action is *maximally* praiseworthy.

Note that in saying that morally worthy actions are maximally praiseworthy I do not intend to say that *different* actions with moral worth cannot warrant different amounts of praise. On the contrary, I will assume that two actions can warrant different amounts of praise due to the difference in their moral significance, even if both actions are morally worthy.<sup>1</sup> For example, it is highly plausible that making tea for your friend is less praiseworthy than saving

---

<sup>1</sup> I hold that rightness does not come in degrees. So, an action's being more or less morally significant than another has nothing to do with it's being more or less morally right than another. What makes an action more or less morally significant than another depends on other factors. I take no stance on what those other factors are, other than that they must correspond to our common-sense intuitions about the moral significance of different types of actions. More work remains to be done here.

a child from a burning building even if both actions are morally worthy. Rather, the point is that an action's having moral worth means that the agent in question deserves the maximum amount of praise possible for performing *that particular* action in that particular context.

Thus, the question is what makes actions maximally praiseworthy in this sense.

Another small detail should be clarified. To say that an action is praiseworthy does not mean that the *action* is praiseworthy per se. Rather, to say that an action is praiseworthy means that the *agent* who performed the action is praiseworthy for performing it. It makes no sense to attribute praise or blame to actions as such. Nevertheless, to keep the language concise, I will often refer to actions as being praiseworthy.

In the literature about moral worth, a common way to frame the dialectic has been that performing morally worthy actions is just 'doing the right thing for the right reasons', where philosophers then debate about what the 'right reasons' are. Here, I will avoid speaking in terms of reasons, but rather in terms of motives. This is partly because of preference. But also because what reasons are, and what it is to act for a reason, is a highly controversial topic.<sup>2</sup> For more on reasons, see Star (2018), and especially part two for issues relevant to this paper. For now, those sympathetic to the reasons-phrasing could think of an agent's motives for acting as his reasons for acting.

Philosophers have generally placed themselves in one of two camps. The first camp holds that what is important for an action's having moral worth is that the agent acts from a moral concern. That is, the agent's object of concern must contain a moral concept. This also naturally includes that the agent has some kind of moral belief. Call this the *Moral Concept View*. For example, what is important for an action's having moral worth might be that it is performed from concern for rightness *as such*. Immanuel Kant (1785 [2018]) is often attributed such a view. It is commonly held that Kant thought an action had moral worth only if it was performed from the motive of duty. Other defenses of the Moral Concept View occur in Sliwa (2016), and Johnson King (2020a).

---

<sup>2</sup> There are also many unclaritys concerning what motivational states are and how they relate to other mental states like beliefs. So, speaking in terms of motives rather than in terms of reasons may not make matters much clearer. Still, I think most have an intuitive idea of what motivational states are. We all know what it *feels* like to be concerned about something and acting because of it. It is less clear to me that we all know what it feels like to have a reason for acting and acting for that reason. For a detailed discussion about the controversies surrounding motivational states, see Schroeder (2020). Here, I simply think of motivational states as mental states with the capacity to move an agent to act.

The second camp instead holds that what is important for an action's having moral worth is that the agent acts from concern for the features that *make* the action morally right. Here, the agent's object of concern will not contain a moral concept. Call this the *Non-Moral Concept View*. For example, what makes it right to donate money to the Red Cross is perhaps the fact that it relieves the suffering of children. Thus, an agent who donates money to the Red Cross because he wants to relieve the suffering of children could perform a morally worthy action even if he does not have any thoughts about acting morally. Some defenses of these views occur in Arpaly (2002), Markovits (2010), Arpaly and Schroeder (2013) and Howard (2021).

There are also alternative views which do not fall into either category. See for instance Tomlinson (2020). However, due to limited space my primary focus will not be on these views.

Some further clarifying remarks about motivational states are needed. There are several possible ways in which an agent could have a concern for the relevant right-making features, or for rightness as such. But the main point of contention has been what it is important that an agent is *non-derivatively* concerned about for an action to have moral worth. It would not be enough, for instance, if an agent cared about the relevant right-making features, or rightness as such, only *because* he had a concern for looking good in the eyes of the public. The reason for focusing on non-derivative concerns is quite simple. It is plain that acting from a direct concern for what is morally relevant warrants more praise than acting from an indirect concern for what is morally relevant.

It is also natural to ask why not morally worthy action requires the agent to have both concern for rightness as such, and concern for the relevant right-making features simultaneously. This is a good question. I can only say that I think such views tend to make morally worthy action too demanding. When your friend gives you a genuine compliment, it is plausible that that action can have moral worth without him being motivated by several things simultaneously (which presumably he is not). One response to my view is simply to insist that morally worthy actions are supposed to be highly demanding and difficult to perform. In any case, these matters must be addressed in detail at another time.

Another important point is that many entries about moral worth implicitly assume objectivism about rightness. This is the idea that what determines whether an action is right are the actual facts of the matter. This stands in contrast with perspectivism about rightness, which is the idea that what determines whether an action is right is (at least partly) the agent's information. Among others, this distinction is examined by Moore (1912 [2005], p. 97-101),

who argued that objectivism is correct. To help illustrate this distinction, I will use a case inspired by Kolodny and Macfarlane (2010), even though they present their case for different purposes. Suppose that a miner is stuck in a mine shaft which is about to crumble, causing his death. There are two buttons in front of you: one blue and one red. Pressing one button will save the miner, while pressing the other will ensure his death. You have no idea which button will save the miner. Fortunately, a highly reliable drone provides you with information that pressing the blue button is most likely to save the miner. You press the blue button. But you are unlucky. In fact, it was the red button which would save the miner. By pressing the blue button, you have ensured his death.

Now one could ask whether you have acted rightly or not. If objectivism is correct, you would have acted wrongly. But if perspectivism is correct, you would have acted rightly. In the literature about moral worth, it is often implicitly assumed that objectivism is correct. This is problematic, as this is a very controversial matter (for the record I prefer perspectivism). Furthermore, I have doubts about if it even makes sense to say that an action could be morally right without it also being praiseworthy.

Since many have thought that an action cannot be morally worthy unless it is also morally right, these considerations undoubtedly affect how one should view moral worth and praiseworthiness in general.<sup>3</sup> Nevertheless, for the purpose of this paper I will ignore these issues. Doing anything else would make the dialectic too complicated and messy, and what I will claim can be adapted to fit either view anyway.

---

<sup>3</sup> There are several possible views to consider here. One could accept that acting rightly is necessary for moral worth in combination with perspectivism. One could also accept that acting rightly is necessary for moral worth in combination with objectivism. Another alternative is to accept either objectivism or perspectivism but deny that acting rightly is necessary for moral worth. Just to take one example, Sliwa (2016) who I will object to in the following sections (sections 3 and 4) thinks that acting rightly is necessary for moral worth. She also accepts objectivism, at least implicitly. So on her view, in failing to press the red button to save the miner you acted wrongly. Because you acted wrongly your action cannot be morally worthy. Hence, the fact that you failed to press the red button diminished your praiseworthiness even though all of your evidence indicated that it was correct to press the blue button. I think this is a very implausible result and it is one of the reasons why I prefer perspectivism – in combination with the idea that acting rightly is necessary for moral worth – to objectivism. An agent's praiseworthiness should not be affected by factors beyond his control. I will note that Sliwa tries to defend her view against considerations like these. I do not think that she succeeds. Nevertheless, I will not pursue these lines of argumentation further in this paper. This is because the ideas I will discuss can be evaluated without committing to either objectivism or perspectivism.

With those clarificatory remarks, I will now turn to argue that the prospects for the Moral Concept View are highly limited.

### 3. Why the Moral Concept View is unsatisfactory: necessary conditions

In the following two sections, I criticize the Moral Concept View by focusing on Paulina Sliwa's (2016) "Moral Worth and Moral Knowledge". Sliwa defends the view that a morally right action is morally worthy just in case the agent acts from concern for doing what is right, accompanied by moral knowledge:

**The Rightness Condition:** A morally right action has moral worth if and only if it is motivated by concern for doing what's right (conative requirement) and by knowledge that it is the right thing to do (knowledge requirement) (Sliwa 2016, p. 394).

I will argue that **The Rightness Condition** provides neither necessary nor sufficient conditions for moral worth. Doing so will reveal that the Moral Concept View is generally limited in two ways. First, acting rightly from a moral concern accompanied by an *explicit* moral belief is not sufficient for moral worth. Second, neither moral concerns nor *explicit* moral beliefs are necessary for moral worth. With respect to moral beliefs, an important detail must be emphasized. I hold that *at least* implicit moral knowledge or moral beliefs are necessary for moral worth, and here I *only* object to the idea that explicit moral knowledge or moral beliefs are necessary for moral worth. If **The Rightness Condition** is interpreted as only requiring at least implicit moral knowledge, it escapes some objections presented here. But more on this later. With that said, let us examine the necessity-part of **The Rightness Condition**.

A much-discussed case threatens to be a counterexample to both the conative requirement and the knowledge requirement as necessary conditions for moral worth. This is the so-called Huckleberry Finn-case, inspired by Mark Twain's novel *The Adventures of Huckleberry Finn* (1884 [2015], Ch. 16). Twain's novel describes a situation where Finn is helping a fugitive slave Jim. During their journey, it strikes Finn that helping Jim is morally wrong, since it occurs to him that "helping a slave is stealing". Finn decides to turn in Jim to the authorities, but changes his mind at the last second, and helps Jim instead. Finn's reasoning is roughly that: "Jim is my friend, so I just

cannot do it". Helping Jim is supposedly the right thing to do. However, Finn helps Jim even though he believes, and continues to believe, that helping Jim is morally wrong.

The correct verdict, as far as I am concerned, is that Finn's helping Jim has, or at least could have moral worth, even though he believes that helping Jim is morally wrong. Finn's action might even seem especially admirable since he had the resolve to act rightly "against all odds". If so, Finn is a counterexample to both the conative requirement and the knowledge requirement as necessary conditions for moral worth.

However, the Huckleberry Finn-case has been the topic of much debate. It has been questioned whether Finn's actions really do have moral worth. This is partly for reasons that will be discussed later. But one reason some might think that Finn's actions cannot have moral worth is because he acts against his own judgement about what he morally ought to do. I will avoid this controversy for now, however. For both the conative requirement and the knowledge requirement as necessary conditions for moral worth fail independently of considerations about Finn. Let us examine some additional counterexamples, starting with the knowledge requirement.

The knowledge requirement as a necessary condition for moral worth fails because of an objection similar to one that Sliwa (2016) discusses. The objection focuses on the fact that sometimes an agent's actions have moral worth even if they have false moral beliefs. Suppose that some version of deontology is the correct moral theory. If the knowledge requirement is a necessary condition for moral worth, that might imply someone's actions could not have moral worth if they believed consequentialism was the correct moral theory. This would be implausible. Even if consequentialism happens to be incorrect, it is still a very reasonable moral theory. Therefore, it is implausible that such a small error would exclude an action from having moral worth.

Sliwa (2016) responds that even if one does not believe in the correct moral theory, that does not exclude one from having moral knowledge. Just as Newtonian physics can grant knowledge about motion even though Newtonian physics is false, a false moral theory can grant moral knowledge even though the moral theory is false. Although consequentialism, deontology, and virtue ethics imply different things, there is still a lot of convergence between them. Having moral knowledge is thus compatible with believing in an incorrect moral theory. This point has also been made by Daniel Star (2008), although in a different context. As Star points out, your friend may firmly believe that his moral knowledge is a result of direct communication with God, whereas you may not believe in God at all. It would be very offensive of your friend to suggest that you completely lack moral knowledge just



because you do not believe in God. It would be equally offensive of you to suggest that your friend completely lacks moral knowledge just because of his religious beliefs. For it should be clear to both of you that, for instance, torturing puppies for fun is wrong.

I think Sliwa's response is successful. But a similar objection works better. Sliwa is correct that believing in some incorrect moral theory is compatible with having moral knowledge. But beliefs in certain metaethical views nevertheless risks excluding moral knowledge. This leads to counterintuitive conclusions.

Suppose Al comes to believe that the error theory is the correct metaethical view. It just so happens that he is incorrect. The correct metaethical view is moral realism. Suppose Al decides to donate money to the Red Cross out of his concern for relieving the suffering of children and his willingness to act according to utilitarian principles. Suppose also that donating money to the Red Cross is the right thing to do. Al holds no inconsistent beliefs, and because he believes the error theory is the correct metaethical view he also actively believes that it is false that it is morally right to donate money to the Red Cross. Hence, Al does not have moral knowledge in this case. The knowledge requirement as a necessary condition for moral worth would imply that Al's action lacks moral worth. But I find it intuitive that Al's action has moral worth.

If you are not yet convinced, suppose Al instead believes that either the error theory or moral realism is correct, but is uncertain about which. Therefore, he suspends judgement about all first-order moral propositions, including the proposition that it is morally right to donate money to the Red Cross. He is a consistent moral skeptic. Nevertheless, Al still finds himself caring about relieving the suffering of children, and also moved by the thought that: if there is a possibility that moral realism is true, or anything matters at all, I want to attempt to act morally. In neither case could Al's actions have moral worth, if the knowledge requirement is a necessary condition for moral worth. This is highly implausible. One's actions can have moral worth even if one is a consistent error theorist or moral skeptic. I conclude that the knowledge requirement as a necessary condition for moral worth is false.

Note that the above cases also threaten the conative requirement as a necessary condition for moral worth. Supposedly, the error theoretician – and perhaps also the moral skeptic – do not act from concern for doing what is right. Given that those agents' actions could nevertheless have moral worth, these cases are also counterexamples to the conative requirement as a necessary condition for moral worth.

The conative requirement as a necessary condition for moral worth also fails for another reason. It is incompatible with the fact that people with a well-developed character

can instinctively act rightly without concern for doing what is right, and yet they perform morally worthy actions. Again consider Al, who discovers that a defenseless woman is being harassed by a large man. Observing this, Al immediately intervenes, even though the size of the man scares him. He intervenes because of his concern for the woman, and because he is courageous enough to overcome his fear of the large man. Al acts instinctively and immediately because of his well-developed character and not because of any concern for doing what is right. In the heat of the moment, it does not occur to him that he is doing what is right or that doing what is right is important. Also suppose that helping the woman is the right thing to do. In this case, I find it intuitive that Al's action has moral worth. But if the conative requirement is a necessary condition for moral worth, Al's action would not have moral worth. So, the conative requirement as a necessary condition for moral worth is false.

Note that in neither case does Al act against his own judgement about what he morally ought to do. This stands in contrast to the Huckleberry Finn-case where Finn does act against his own judgement about what he morally ought to do. Those who think Finn's actions cannot have moral worth because he acts against his own judgement about what he morally ought to do should be more impressed by the Al-cases.

Neither the conative requirement nor the knowledge requirement are necessary conditions for moral worth. In conclusion, neither moral concerns, nor explicit moral beliefs are necessary for moral worth.

#### **4. Why the Moral Concept View is unsatisfactory: sufficient conditions**

Can **The Rightness Condition** instead provide a sufficient condition for moral worth? Or in other words, are the knowledge requirement and the conative requirement jointly sufficient for moral worth? Here, I will consider the objection inspired by Michael Smith (1994), who argues that concerns for rightness *as such* are morally problematic. Assuming that Smith is correct it is safe to say that: because concerns for rightness as such are morally problematic, actions performed *only* from concern for rightness as such – or any other moral concern – are not morally worthy. Consequently, **The Rightness Condition** as a sufficient condition for moral worth would be false. The question is thus whether or not concerns for rightness as such are morally problematic, and it is the question that will be treated for the remainder of this section.

The idea that concerns for rightness as such are morally problematic can be supported by a well-known case from Bernard Williams (1981). Suppose you discover that

your wife is drowning. In the heat of the moment, you become motivated to save your wife, and do so. You do not save your wife because of any concern for her, but because of your concern for doing what is right accompanied by the knowledge that saving your wife is the right thing to do. **The Rightness Condition** as a sufficient condition for moral worth implies that saving your wife had moral worth, since your action was motivated by concern for doing what is right, accompanied by the knowledge that your action was right. But to be motivated in the heat of the moment to save your wife from drowning, only because of your concern for doing what is right without any direct concern for your wife herself, is a motivation which seems morally problematic. As mentioned above, Smith (1994) similarly suggests that concerns for rightness as such are morally problematic, and that therefore it is repugnant to think that morally exemplary agents are characterized by such concerns.

Williams and Smith present their arguments in slightly different contexts, and both arguments have come under much criticism. See for instance Lillehammer (1997) and Olson (2002) for such criticism (and Zhang (2021) for a recent defense of Williams and Smith). But what cases such as Williams' show is at least that an agent who is *only* concerned with doing what is right without concern for anything else is not a morally exemplary agent. If that is correct, then cases such as Williams' are counterexamples to **The Rightness Condition** as a sufficient condition for moral worth.

Sliwa responds by giving two reasons why a concern for rightness as such is not morally problematic, and why it is "...an indispensable part of the motivational set of a morally good person." (Sliwa 2016, p. 408). First, she thinks concerns for doing what is right are essential when faced with moral uncertainty. When uncertain about what the right thing to do is, the morally good agent will try to find out what that is. And to do so, she must presumably be motivated to do what is right. Second, the morally good agent is disposed to experience a wide variety of emotions such as remorse and moral indignation. And to experience such emotions, she must presumably conceptualize her actions in moral terms. That is, conceptualizing them as right or wrong, and also be concerned with doing what is right.<sup>4</sup>

Regardless of the correctness of Sliwa's claims above, they are irrelevant in the current context. Even if it is true that a concern for doing what is right is an important

---

<sup>4</sup> Sliwa (2016, p. 406-408) addresses these matters when defending the conative requirement as a necessary condition for moral worth. This is somewhat confusing as the charge from Williams and Smith better supports an objection to **The Rightness Condition** as a sufficient condition for moral worth.

*component* of being a morally exemplary agent, it is still the case that someone who is *only* concerned with doing what is right without concern for anything else cannot be a morally exemplary agent. That is all it takes for the arguments above to work as an objection against **The Rightness Condition** as a sufficient condition for moral worth.

Johnson King (2020b) responds to Williams and Smith in a different way. She argues that by considering certain cases, we can see that there is nothing morally problematic about concerns for rightness as such. She demonstrates by letting the reader compare two cases, which she calls ‘a minimal pair’. I will only consider the first case. This is not to mislead the reader, but only because considering the first case will suffice for my purposes. Here it is:

Maryam is chairing a session at a prestigious Philosophy conference, which is notorious for getting nasty during Q&A. Maryam wants to act rightly – that is, she wants to conduct Q&A in such a manner as to meet all of her obligations not only *qua* chair but also *qua* moral agent. So she thinks carefully about what her obligations might be, planning to modify her behavior in light of her conclusions. After much soul-searching and subtle reasoning, Maryam decides that four things matter morally in her case: prioritizing junior scholars over senior scholars, prioritizing those who have asked fewer questions at the conference over those who have asked lots already, discouraging audience members from asking repeated versions of the same question, and discouraging them from battering the speaker with multiple lengthy follow-ups. Maryam devises a set of principles that allows her to promote these four ends in a manner that reflects her estimation of their relative importance. She then conducts Q&A in perfect accordance with her principles. Moreover, Maryam is completely right about all of this. She has exhaustively identified the things that matter morally under her circumstances, and has chosen principles that precisely reflect their relative importance. Maryam has perfected the novel field of Conference Ethics. Since she guides her behavior in accordance with her conclusions, she acts perfectly (Johnson King 2020b, p. 415-416).

Here, Maryam’s actions are ultimately driven by concern for doing what is right. Johnson King (2020b) argues that it is highly implausible that Maryam is not praiseworthy at all. And if it is highly implausible that Maryam is not praiseworthy at all, we should doubt that there is something morally problematic about concerns for doing what is right.

Johnston King is correct that it is highly implausible that Maryam is not praiseworthy at all. So, Johnson King succeeds in showing that, at least in some cases, one is worthy of praise even if one acts from concern for doing what is right. This is to be expected. Saving your drowning wife because you are concerned about doing what is right certainly seems better, and more praiseworthy, than saving your wife because you are afraid that not doing so would hurt your reputation.

However, Johnson King (2020b) also thinks that Maryam's motivational disposition is in no way morally sub-optimal. This is implausible. From the fact that Maryam is praiseworthy to some degree it does not follow that Maryam is maximally praiseworthy in the way required for moral worth. At face value Maryam might seem like a moral saint. Maryam acts flawlessly from concern for doing what is right and has engaged in flawless moral reasoning, which is good. But upon further inspection, the facade falls apart. Remember that we are being asked to imagine that Maryam has no direct concern whatsoever for the actual subjects of moral importance, such as junior scholars. True, Maryam has a derivative concern for the junior scholars. But were it not for Maryam's concern for doing what is right, she would not care whatsoever about the junior scholars. Once again, this seems rather cold. Compare with Williams' wife-saver: would Williams' wife-saver really seem that much better if we emphasized that he cared about his wife, but only because he wanted to do what is right?

So, although Maryam's actions are praiseworthy to some degree, they are not morally worthy. For the morally exemplary agent should care directly about the things of moral importance. So, the objection inspired by Williams and Smith still stands as a strong objection to **The Rightness Condition** as a sufficient condition for moral worth. In conclusion, acting rightly from a moral concern accompanied by an explicit moral belief is not sufficient for moral worth.

I have now argued that the prospects for the Moral Concept View are highly limited. Unfortunately the supposed alternative, the Non-Moral Concept View, might be an even worse option. For it is subject to a serious problem that does not befall the Moral Concept View. I will turn to examine this next.

## **5. The Non-Moral Concept View and accidentally doing the right thing**

The previous discussion indicates that the Moral Concept View is problematic. This potentially makes the Non-Moral Concept View look more attractive. However, adherents of

the Moral Concept View have objected to the Non-Moral Concept View by arguing that unless an agent acts rightly from concern for rightness as such, or anything similar, that agent would accidentally act rightly, in the sense that his action would merely coincidentally coincide with moral demands. And presumably, accidentally right actions cannot have moral worth. In this section I clarify this objection, and object to two proposed solutions. This will help motivate my own account which will be developed in section 7.

Let me begin by saying that there are several senses in which someone could accidentally act rightly. But not all senses of ‘accidental’ affect the praiseworthiness of an action. Suppose Harold is a perfect moral agent, but also unable to move. He can only use his superb moral skills for one day if lightning strikes him, temporarily relieving him of his paralysis. Luckily, Harold is struck by lightning and acts rightly for one day as a consequence.

There is a sense in which it is an accident that Harold acts rightly. But this type of accident does not affect the praiseworthiness of his actions. Presumably, the accidental features of the background conditions that allow agents to act rightly does not speak ill or well of their actions. The sense of ‘accidental’ that can negatively affect the praiseworthiness of an action must instead have something to do with the agent himself. It must have something to do with how the agent is disposed, both emotionally and cognitively. This is what the forthcoming objection to the Non-Moral Concept View targets.

The objection is made clear by considering the simplest version of the Non-Moral Concept View:

**The Simple View:** An action has moral worth if, and only if, (1) the action is morally right, and (2) the action is performed from concern for the relevant right-making features.

**The Simple View** avoids the problems that befall the Moral Concept View discussed in the previous sections. But critics complain that **The Simple View** entails that morally worthy actions are, or at least could be, instances of people accidentally acting rightly. And that would be unacceptable. There are several ways to present the objection. But the most intuitive way to understand it is by considering the following:

Consider Bob, who decides to give a bystander directions because he is motivated to save the bystander some time, and he knows that giving the bystander directions is a

way of doing so. Suppose that giving the bystander directions is the right thing to do, and the fact that it will save the bystander some time is the relevant right-making feature. However, Bob is disposed such that no matter what situation he is in, his motivation to save bystanders some time is the dominant one.

Here Bob acts rightly because of his concern for the relevant right-making feature: that his action will save the bystander some time. According to **The Simple View**, Bob's action would thus have moral worth. But ascribing moral worth to Bob's action is dubious. It is difficult to pinpoint exactly what is problematic about Bob. But roughly, the problem is that Bob's motivation to save the bystander some time seems to not be dependent whatsoever on the fact that it is morally appropriate to do so in the situation Bob is in. For Bob is disposed such that he would have prioritized saving bystanders some time even if it was not morally appropriate to do so. Thus, it seems Bob accidentally acts rightly, in the sense that his action merely coincidentally coincides with moral demands. And only non-accidentally right actions can have moral worth.

By contrast, there are no similar problems with agents who act from concern for doing what is right. For their motivation to act in a certain way does depend on whether it is morally appropriate to act in that way in the situation they are in. So, it is easier to justify that they non-accidentally act rightly:

Consider Carl, who decides to give a bystander directions because he is motivated to do what is right, and he knows that giving the bystander directions is the right thing to do. Furthermore, Carl is disposed such that no matter what situation he is in, his motivation to do what is right is the dominant one.

So, in this respect the Moral Concept View has the advantage over the Non-Moral Concept View. How should an adherent of the Non-Moral Concept View respond? I will review two suggestions and find them both wanting.

Markovits (2010) argues that despite appearances, someone like Bob non-accidentally acts rightly. Thus, we can accept the conclusion that Bob's action has moral worth. She asks us to consider a fanatical dog-lover who heroically performs a dangerous rescue operation out of concern for some strangers who are in danger. It seems unreasonable to suggest that the hero's actions would be less praiseworthy just because it is true that *if* a dog happened to also be in danger, he would have saved the dog instead of the strangers.

Markovits argues that cases like these show that what an agent *would* have been motivated by, and done, in counterfactual situations does not affect the praiseworthiness of his *actual* motives and actions. Hence, the fact that Bob's motivational disposition would lead him to act wrongly in various counterfactual situations does not affect the praiseworthiness of his actual motives and actions either. Markovits concludes that acting rightly from concern for the relevant right-making features is sufficient for non-accidentally acting rightly, regardless of what the agent would have done in counterfactual situations.

But this final claim is implausible. For Markovits' defense of **The Simple View** to work, it must be case that acting rightly from concern for the relevant right-making features is sufficient for non-accidentally acting rightly simply because there is a metaphysical 'make it the case-relation' between the relevant right-making features and the rightness of the action. But as Johnson King (2020a) notes, this understanding of 'non-accidental' does not generalize well to other areas. Understanding 'non-accidental' in this way seems to imply that:

...for someone to non-accidentally perform an act of type A it is sufficient that (a) she is motivated to perform it by the fact that it is of type B and (b) as a matter of metaphysical fact, the agent's performing an act of type B makes it the case that she performs an act of type A (Johnson King 2020a, p. 193).

Johnson King (2020a) borrows various examples to demonstrate how this leads to counterintuitive conclusions. Here is one from Wayne D. Riggs (2014):

Emilia has been running away from a gang of vampires throughout the night. They have chased her all through her home town. They have been so close on her heels that she has been unable to stop to ask anyone for help. She eventually reaches the edge of town and begins to run across an empty field. Finally, though, she reaches her limit and stumbles exhausted to the ground. Inevitably, the vampires catch up to Emilia and immediately go in for the kill. Just at that moment, though, the sun peeks over the horizon, catching all the vampires in the open. They all burst into flame (as vampires must do at the touch of the sun's rays), and Emilia is saved. She gasps in unexpected relief (Riggs 2014, p. 633).



In this case, there is a metaphysical ‘make it the case-relation’ between running into the open fields and luring the vampires to their deaths, in the sense that running into the open fields is what makes it the case that Emilia lures the vampires to their deaths. Emilia is motivated to run into the open fields. But it seems obvious that Emilia merely accidentally lures the vampires to their deaths. So, luring vampires to their deaths from concern for the features that make it the case that the act is an instance of luring vampires to their deaths does not guarantee that one non-accidentally lures vampires to their deaths. Similarly, acting rightly from concern for the features that make it the case that the act is right does not guarantee that one non-accidentally acts rightly.

So, Markovits’ insistence that someone like Bob non-accidentally acts rightly is highly implausible. However, it is still noteworthy that intuitively, Markovits’ dog-lover non-accidentally acts rightly even though he, like Bob, acts from concern for the relevant right-making features without any concern for rightness as such. This suggests that there are subtleties which are missing from the discussion.

This leads to the second suggestion, which comes from Nathan Howard (2021). Howard would accept that someone like Bob accidentally acts rightly and that therefore his action does not have moral worth. But that is only because Bob’s motivational disposition is such that he would have prioritized saving bystanders some time regardless of what situation he was in. Not all people are like this. Many people who act rightly from concern for the relevant right-making features are disposed to prioritize differently if the moral situation is different. In contrast with Bob, those people are disposed such that they could not easily have acted wrongly. Hence, it would be reasonable to assume that people who act rightly from concern for the relevant right-making features non-accidentally act rightly, given that they are also sensitive to changes in their moral situation. If so, their actions could have moral worth even if Bob’s do not.

To illustrate, Howard (2021) considers a soccer player named Renaldo, who is simply amazing at soccer. His unparalleled talent and impeccable sense of the game makes him untouchable on the field. Every time he plays one can expect to witness athletic superiority in action. However, Renaldo is also extremely humble, and has huge standards for what would characterize a great play. He does not believe that his plays are great. Neither does he have any concern for making great plays. When asked why he plays soccer and what motivates him he simply answers that: “I like soccer and winning, so I just try to place the ball in the net and help my teammates do so”.

Even though Renaldo does not believe that his plays are great and has no concern for making great plays, it is no accident that he makes great plays. His great plays are the result of countless hours of practice on the soccer-field. His great plays are non-accidental, because he is disposed to prioritize differently depending on the state of the game. Compare Renaldo with an amateur who does not believe that his plays are great and has no concern for making great plays, but randomly kicks the ball in a desperate attempt to be useful to his team and manages to score a miraculous goal. The amateur only accidentally makes a great play, because he is not disposed to prioritize differently depending on the state of the game. If the ball had been in a slightly different position, his desperate attempt would have resulted in a bad play. This is not the case with Renaldo. So, Renaldo can non-accidentally play greatly whereas the amateur accidentally plays greatly, even though none of them have any thoughts about making great plays.

Howard's (2021) idea is that instances of non-accidental athletic performance and non-accidental moral performance are analogous in this respect. Howard does not say this explicitly, but what he is essentially doing is suggesting the following modification (or something similar) to **The Simple View**:

**The Modal View:** An action has moral worth if, and only if, (1) the action is morally right, (2) the action is performed from concern for the relevant right-making features, and (3) the agent acts from a disposition that will not easily lead him to act wrongly.

Here conditions (1) and (2) are the same as in **The Simple View**, and the addition of (3) is supposed to ensure that actions with moral worth are not instances of people accidentally acting rightly.

But questions remain. Sliwa (2016) has pointed out that if we accept something like **The Modal View**, it seems we must also accept that how praiseworthy an agent is for an action depends on what they *would* have done in various counterfactual situations. But that is counterintuitive. In general, it seems that how praiseworthy an agent is for an action depends only on what they *actually* do, and not on what they would do if the situation was different (recall Markovits' dog-lover).<sup>5</sup> Furthermore, Sliwa points out that accepting **The Modal View** would leave us with a technical difficulty concerning where to draw the line with

---

<sup>5</sup> Although some accept that counterfactual situations affect praiseworthiness. See for instance Zimmerman (2002) for such a view.

respect to exactly how many counterfactual situations an agent must have been motivated properly and acted rightly in, in order for them to count as having non-accidentally acted rightly. If there is no non-arbitrary line which can be drawn, an account of moral worth which relies on what the agent would have been motivated by, and done, in various counterfactual situations risks being unclear. One could respond that it is simply vague where this line should be drawn. I am unsure about if this is a good response or not. In any case, these considerations cast doubt on Howard's proposal.

My diagnosis is that there is an underlying reason these attempts at rescuing the Non-Moral Concept View fail. They both fail because of a misunderstanding of what it takes to act rightly non-accidentally. Clarifying this issue will reveal the proper way forward. This is where I turn next.

## **6. What it takes to act rightly non-accidentally**

The challenge set out for adherents of the Non-Moral Concept View is to explain how agents can non-accidentally act rightly in a way that is compatible with the Non-Moral Concept View. In particular, adherents of the Non-Moral Concept View must explain how agents can non-accidentally act rightly *without any thoughts about acting morally*. But so far, there has been a misunderstanding of what it takes to act rightly non-accidentally. People like Markovits' dog-lover and Howard's Renaldo are indeed different from Bob in the sense that it is plausible that the former non-accidentally perform their respective actions, whereas Bob does not. But a misunderstanding of what it takes to act rightly non-accidentally has prevented anyone from identifying what the relevant difference between them could be.

In what follows, my focus will be on remedying this mistake and clarifying what it takes to act rightly non-accidentally. My primary aim will not be to give a complete analysis of what it takes to act rightly non-accidentally (although the discussion will hint at such an analysis). Such a task would exceed the scope of this paper. Since my intent is to examine whether there is any way to defend the Non-Moral Concept View, my primary aim is instead to try to find a sufficient condition for non-accidentally acting rightly that can explain how agents can non-accidentally act rightly in a way that is compatible with the Non-Moral Concept View. If at least that is possible, there is hope for the Non-Moral Concept View.

One attempt at giving a sufficient condition for non-accidentally acting rightly is to spell it out in counterfactual terms:

**Counterfactual Robustness:** An agent non-accidentally acts rightly if the agent's acting rightly is counterfactually robust, in the sense that the agent acts from a set of dispositions that will not easily lead him to act wrongly.

Attempting to explain how agents can non-accidentally act rightly in a way that is compatible with the Non-Moral Concept View by invoking a counterfactual condition mirrors Howard's (2021) approach. However, the use of counterfactual conditions is not ideal in this context. It may be that some counterfactual conditions are necessary conditions for non-accidentally acting rightly, and thus may give satisfying explanations of why someone like Bob accidentally acts rightly. However, they are not sufficient conditions for non-accidentally acting rightly.

This can be shown with some help from the epistemology literature, where various counterfactual conditions such as sensitivity and safety have been proposed as anti-luck conditions on knowledge in an attempt to solve Gettier-problems (Gettier 1963). See for example Nozick (1981), Sosa (1999), Williamson (2000), and Pritchard (2005). It has been shown convincingly that counterfactual conditions fail to exclude luck in the way required for knowledge. Jennifer Lackey (2006) makes this point:

The fundamental problem with such modal accounts is that counterfactual robustness can be ensured through a combination of features that is entirely fortuitous. ...because modal accounts lack the resources to ensure a proper connection between mind and world, counterfactual robustness fails to be a reliable guide to the presence of either luck or knowledge (Lackey 2006, p 289).

Counterfactual conditions for non-accidentally acting rightly fail for similar reasons. Suppose that there was a machine X, whose only function was to give the correct moral advice in every situation. Now suppose an unfortunate subject Denise has been born with an irresistible and obsessive urge to follow the advice of machines of type X no matter what situation she is in. Suppose also that machine X is indestructible, and that therefore Denise has access to machine X in every counterfactual situation. Denise has no idea that machine X will give the correct moral advice in every situation. Denise acts rightly as a consequence.

Here Denise's acting rightly fulfills **Counterfactual Robustness**. For acting from a disposition to do whatever machines of type X say will never lead Denise to act wrongly regardless of what situation she is in. This makes Denise's acting rightly perfectly

counterfactually robust. But it still seems that Denise accidentally acts rightly, as her decision to act in accordance with machine X's advice has nothing to do with morality.

This is a general problem with all sufficient conditions for non-accidentally acting rightly which are spelled out in purely counterfactual terms. No counterfactual condition by itself is a sufficient condition for non-accidentally acting rightly. Hence, a purely counterfactual condition cannot explain how agents can non-accidentally act rightly in a way that is compatible with the Non-Moral Concept View. For counterfactual conditions cannot explain how agents can non-accidentally act rightly at all.

Another attempt at giving a sufficient condition for non-accidentally acting rightly is to spell it out in non-counterfactual terms. Here are two attempts:

**Make It the Case:** An agent non-accidentally acts rightly if the agent acts rightly from concern for the relevant right-making features.

**Deliberately:** An agent non-accidentally acts rightly if the agent deliberately does the right thing (see Johnson King (2020a) for a view along these lines).

I covered the problem with **Make It the Case** in the previous section. It is highly implausible that **Make It the Case** is a sufficient condition for non-accidentally acting rightly, since this understanding of 'non-accidental' does not generalize well to other areas.

What about **Deliberately**? Contrary to counterfactual conditions, **Deliberately** (and **Make It the Case**) avoids trouble with machine X. For it is obvious that Denise does not deliberately act rightly when she follows machine X's advice. This is because deliberately acting rightly at least implies that one intends to perform the action under the description of it being the right thing to do, and Denise does not. In the absence of further counterexamples, **Deliberately** indeed seems like a sufficient condition for non-accidentally acting rightly. Unfortunately, adherents of the Non-Moral Concept View have no use for **Deliberately**. For as mentioned above, deliberately acting rightly at least implies that one intends to perform the action under the description of it being the right thing to do. So, although **Deliberately** can explain how agents can non-accidentally act rightly, it cannot explain it in a way that is compatible with the Non-Moral Concept View. For adherents of the Non-Moral Concept View must explain how agents can non-accidentally act rightly without any thoughts about acting morally.

So far, things are looking dim for the Non-Moral Concept View. The failures so far reveal the difficulty of explaining how agents can non-accidentally act rightly without *any* thoughts about acting morally. In fact, I must concede that this project looks hopeless. To act rightly non-accidentally, it seems the agent must have made use of his own abilities to interact with his moral reality in some way or another. As Johnson King says: “For all types of acts A, someone accidentally As if she has no idea that she is performing an act of type A when she does so.” (Johnson King 2020a, p. 196). I agree – an agent cannot non-accidentally act rightly without any thoughts about acting morally.

So, as presently understood the Non-Moral Concept View is untenable. To rescue the Non-Moral Concept View, the claim that a morally right action can be non-accidentally right and have moral worth without the agent having any thoughts about acting morally must be weakened. This claim has so far been interpreted as:

**The Strong Claim:** A morally right action can be non-accidentally right and have moral worth without the agent’s having any thoughts *whatsoever* about acting morally.

This claim is too strong, leaving adherents of the Non-Moral Concept View in a hopeless position. Instead, adherents of the Non-Moral Concept View should give some ground and accept:

**The Weak Claim:** A morally right action can be non-accidentally right and have moral worth without the agent’s having any *explicit* thoughts about acting morally.

The difference between implicit and explicit thoughts is that to have an explicit thought about some fact, the thought must have reached the agent’s conscious awareness. An implicit thought about some fact, on the other hand, does not require this. Here it is instead enough that the agent has an unconscious thought about that fact, or a bodily awareness of that fact.

As the Non-Moral Concept View has been presented so far, it has involved accepting **The Strong Claim**. So, let us therefore call the version of the Non-Moral Concept View which involves denying **The Strong Claim** in favor of **The Weak Claim** the *Weak Non-Moral Concept View* as opposed to the *Strong Non-Moral Concept View*. While the Strong Non-Moral Concept View is untenable, I will now argue that the Weak Non-Moral Concept View is not.

While **The Strong Claim** is implausible, **The Weak Claim** is not. For it is false that one non-accidentally performs an action of a certain type, only if one has an explicit thought about performing an action of that type. Suppose Earl is walking in the jungle and hears a rustle in the grass ahead of him. Earl explicitly believes (firmly) that the rustle is nothing to worry about, but he feels uneasy about it and decides to walk around the rustle-area anyway. In doing so, Earl avoided a venomous snake which was causing the rustle. Earl has avoided danger.

Here Earl does not have any explicit thoughts about avoiding danger. However, it still seems that Earl non-accidentally avoids danger (Earl does not just coincidentally avoid danger). For even though Earl does not have any explicit thoughts about avoiding danger, he has still *implicitly* recognized the danger that the rustle indicates. His implicit recognition of the danger that the rustle indicated triggered an automatic emotional response: a fear of the rustle-area, which motivated him to walk around the rustle-area. This suffices for saying that Earl non-accidentally avoids danger, and that he can be credited for avoiding danger.

A sufficient condition for non-accidentally acting rightly, which can explain how agents can non-accidentally act rightly in a way that is compatible with the Weak Non-Moral Concept View, can be found by examining why Earl non-accidentally avoids danger. The best explanation why Earl non-accidentally avoids danger, even though he does not have any explicit thoughts about avoiding danger, is that his motivation to act in a way that avoids danger is nevertheless *explained* by his recognition of the danger. I suggest an analogous sufficient condition for non-accidentally acting rightly:

**Explained:** An agent non-accidentally acts rightly if the agent's motivation to perform the action is *explained* by his recognition of the relevant moral facts.

Note that to recognize the relevant moral facts, the agent must have made use of relevant moral concepts: moral rightness, moral goodness, or normative forcefulness etc. So, even if some synthetic naturalism happened to be correct and the property of rightness is identical to the property of, say, maximizing happiness it would not be enough for an agent to recognize that some action maximizes happiness for him to have recognized the relevant moral facts. What is intended is instead that the agent must have recognized the relevant moral facts by making use of relevant moral concepts.

In any case, while **Explained** cannot aid the Strong Non-Moral Concept View, it can explain how agents can non-accidentally act rightly in a way that is compatible with the Weak Non-Moral Concept View.

## 7. An account of moral worth

By focusing on **Explained**, we are well-equipped to give an account of non-accidentally acting rightly, and by extension, moral worth. In fact, **Explained** is embedded in my preferred account of moral worth:

**The Explanation View:** An action has moral worth if, and only if, (1) the action is morally right, (2) the action is performed from concern for the relevant right-making features, and (3) the agent's motivation to perform the action is explained by his recognition of the relevant moral facts.

Here condition (1) and (2) are the same as in **The Simple View**, and the addition of (3) ensures that actions with moral worth are not instances of people accidentally acting rightly. I will now examine the implications of **The Explanation View**, and then address some questions and objections.

By proposing that a similar emotional function as with Earl is often at work regarding moral matters, **Explained** can explain how agents can non-accidentally act rightly in a way that is compatible with the Weak Non-Moral Concept View. The suggestion is that similarly to our ability to recognize implicitly that features of the world indicate or are *dangerous*, thereby triggering an automatic negative emotional response, we have an ability to recognize implicitly the *normative force* of features in the world (recognizing the relevant moral facts), thereby triggering an automatic emotional response. When agents implicitly recognize the normative force of features in the world, their emotional systems automatically trigger a *non-derivative* concern for those features.

For example, suppose Fatima comes across an old lady who has dropped her glasses and cannot find them. Fatima notices that the old lady is in distress, and she immediately becomes concerned about relieving her distress. As a consequence, Fatima helps the old lady find her glasses. Suppose that helping the old lady find her glasses is the right thing to do, and the fact that it will relieve her distress is the relevant right-making feature.



The diagnosis of this situation could now be similar to Earl and the danger-example. Fatima has automatically become concerned about relieving the old lady's distress, as a result of implicitly recognizing the normative force of the old lady's distress. Consequently, Fatima acts rightly from concern for the relevant right-making features. Similar to Earl's not having any explicit thoughts about avoiding danger, Fatima does not have any explicit thoughts about acting morally either. Nevertheless, Fatima still non-accidentally acts rightly, as her motivation to help the old lady find her glasses is explained by her recognition of the relevant moral facts. It is explained by her recognition of the relevant moral facts because Fatima's concern for relieving the old lady of her distress is an automatic result of her implicit recognition of the normative force of the old lady's distress. Hence, Fatima's action has moral worth according to **The Explanation View**.

In summary, here is the description of the typical morally worthy action. An agent implicitly recognizes the normative force of some right-making features. This triggers an automatic emotional response: a non-derivative concern for those features. The agent then performs a morally right action from that concern. In doing so, the agent's action will satisfy the following conditions: (1) the action is morally right, (2) the action is performed from concern for the relevant right-making features, and (3) the agent's motivation to perform the action is explained by his recognition of the relevant moral facts, ensuring that the agent non-accidentally acts rightly. Hence, that action has moral worth according to **The Explanation View**.

The upshots of **The Explanation View** are that we can maintain that morally worthy actions must be performed from concern for the relevant right-making features, while simultaneously ensuring that all morally worthy actions are instances of people non-accidentally acting rightly. It also avoids the problems with the Moral Concept View discussed in the previous sections. So, we have good reasons to accept it. As a consequence, one must admit that **The Strong Claim** is false, and instead endorse **The Weak Claim**. But making this concession should not worry adherents of the Non-Moral Concept View. All intuitions about praiseworthiness that makes the Non-Moral Concept View attractive can be accounted for without **The Strong Claim**. To demonstrate, let us consider two previously discussed cases in light of **The Explanation View**.

Consider first the Huckleberry Finn-case, which was briefly discussed in section 3. The case involves Huck helping a fugitive slave Jim even though Finn believes that doing so is morally wrong. The diagnosis of what is going on with Finn is that although Finn explicitly believes that helping Jim is morally wrong, Finn implicitly recognizes the normative force of

the fact that Jim is his friend. This triggers an automatic emotional response: a non-derivative concern for their friendship. Because of this, Finn is motivated to help Jim and does so. Finn thereby acts rightly from concern for the relevant right-making feature (the fact that Jim is his friend), and Finn's motivation to perform the action is explained by his recognition of the relevant moral facts (the normative force of the fact that Jim is his friend). Hence, **The Explanation View** yields the correct verdict that Finn's helping Jim has moral worth. The case of Huckleberry Finn demonstrates how our explicit and implicit beliefs can conflict, and how this can lead to hesitant behavior and feelings of internal conflict.

Also consider the case of Markovits' dog-lover, which was briefly discussed in section 5. The case involves a fanatical dog-lover who heroically rescues some strangers who are in danger. But *if* a dog had happened to also be in danger, he would have saved the dog instead of the strangers. The diagnosis of what is going on with the dog-lover is that he implicitly recognizes the normative force of the strangers' lives. This triggers an automatic emotional response: a non-derivative concern for the strangers' lives. Because of this, the dog-lover is motivated to rescue the strangers and does so. The dog-lover thereby acts rightly from concern for the relevant right-making feature (the strangers' lives), and the dog-lover's motivation to perform the action is explained by his recognition of the relevant moral facts (the normative force of the strangers' lives). Hence, **The Explanation View** yields the correct verdict that the dog-lover's saving the strangers has moral worth. Of course, if a dog happened to also be in danger, the dog-lover's obsession would no longer remain dormant, and would then interfere with his heroic motives. But since no dog is in danger in the actual case, the dog-lover's heroic motives remain untouched. Thus, **The Explanation View** neatly explains why the dog-lover's counterfactual motives and actions do not affect the praiseworthiness of his actual motives and actions.

Of course, it is not required in either the case of Huckleberry Finn or the dog-lover that we *must* conceive of the cases as described above. It is certainly possible to conceive of the possibility that Finn's concern for his friendship with Jim and the dog-lover's concern for the strangers' lives are not explained by their recognition of the relevant moral facts. If the cases are conceived of in this manner, neither Finn's nor the dog-lover's actions would have moral worth according to **The Explanation View**. Indeed, if Finn and the dog-lover's motivations to perform their respective actions was not explained by their recognition of the relevant moral facts, then their acting rightly would be no less accidental than Denise's blindly following the advice of machines of type X. The fact that these cases can be conceived of in these different ways is a plausible reason why adherents of the Non-Moral

Concept View and the Moral Concept View have had such differing intuitions about them. When an adherent of the Non-Moral Concept View has conceived these cases, he might have (unconsciously) conceived of the agent's motives to act as being explained by his recognition of the relevant moral facts. When an adherent of the Moral Concept View has conceived these cases, he might instead have conceived of the agent's motives to act as *not* being explained by his recognition of the relevant moral facts.

It should now be plain that there are plenty of reasons to accept **The Explanation View**. Before concluding, I will respond to some questions and objections.

## 8. Questions and objections

First, a point of clarification. I have argued that morally worthy action requires the agent to have at least an implicit moral belief. This retrospectively affects some arguments presented in section 3. In particular, it affects my objections to the knowledge requirement as a necessary condition for moral worth. I argued that the knowledge requirement as a necessary condition for moral worth is false because it excludes the actions of a consistent error theorist and moral sceptic from having moral worth. This is because for any first-order moral proposition, a consistent error theorist would believe that proposition to be false and a consistent moral sceptic would suspend judgement about it. Hence, neither the consistent error theorist nor moral sceptic could have moral knowledge. The knowledge requirement as a necessary condition for moral worth thus implies that their actions cannot have moral worth. This would be implausible.

However, it is possible to defend a version of the knowledge requirement as a necessary condition for moral worth by making a distinction between explicit and implicit moral knowledge. Even if the consistent error theorist and moral sceptic do not explicitly believe any first-order moral propositions, they may still implicitly believe many first-order moral propositions. Hence, it is possible to claim that even though the consistent error theorist and moral sceptic do not have explicit moral knowledge, they may still have a great deal of implicit moral knowledge. If in turn the knowledge requirement as a necessary condition for moral worth is interpreted as only requiring implicit moral knowledge, it would not exclude the actions of a consistent error theorist or moral sceptic from having moral worth. Hence, neither the consistent error theorist nor moral sceptic would work as counterexamples to the knowledge requirement as a necessary condition for moral worth, if the requirement is interpreted as only requiring implicit moral knowledge. I can only

speculate about how the originator of the knowledge requirement – Sliwa – would want it to be interpreted. My guess is that the possible distinction between explicit and implicit moral knowledge is not something she has considered.

That being said, the consistent error theorist and moral sceptic are still counterexamples to the knowledge requirement as a necessary condition for moral worth, if the requirement is interpreted as requiring explicit moral knowledge. This is an important conclusion in itself. Furthermore, no matter how the knowledge requirement is interpreted **The Rightness Condition** as a whole still fails because of the shortcomings of the conative requirement.

I will now turn to address some questions and objections. One worry one might have is that I have only stated **Explained** as a *sufficient* condition for non-accidentally acting rightly. But if **Explained** is not also a necessary condition for non-accidentally acting rightly, this leaves it possible that there could be other ways of non-accidentally acting rightly in a way that is compatible with the Weak (or Strong) Non-Moral Concept View. If so, then **The Explanation View** risks excluding actions which are morally worthy and would at most work as a sufficient condition for moral worth. This worry is understandable, and it remains to be seen whether **Explained** is also a necessary condition for non-accidentally acting rightly. However, I see no good reason to think that it is not. To show that **Explained** is not a necessary condition for non-accidentally acting rightly we would need a case where an agent acts rightly from concern for the relevant right-making features, his motivation to perform the action is *not* explained by his recognition of the relevant moral facts, and the agent still non-accidentally acts rightly. I cannot think of any such cases. Nevertheless, I recognize that more work needs to be done here.

Another natural question to ask is if accepting **The Explanation View** also requires one to accept some kind of motivational internalism. After all, in describing how **The Explanation View** works in practice, I claimed that the typical morally worthy action involves that the agent's concern for the relevant right-making features is the *automatic* result of his implicit recognition of the normative force of those features. This resembles a kind of motivational internalism, according to which making the implicit judgement that some feature has normative force must include having some degree of non-derivative concern for that feature. Strictly speaking the answer is no: accepting **The Explanation View** does not require one to also accept any motivational internalism.

However, accepting **The Explanation View** in conjunction with this kind of internalism has an advantage to not doing so, namely that it makes the job of accounting for

the explanatory link between one's recognizing the relevant moral facts, one's subsequent concern for the relevant right-making features, and one's subsequent motivation to perform a particular action that is morally right much easier. I favor this kind of internalism for this reason, and you should too.

But some might be skeptical. A common reason to reject motivational internalism is the apparent conceptual possibility or even actual existence of so-called amoralists. That is, people who seem to make moral judgements, but remain completely emotionally unaffected by them. Consider the psychopath who agrees that his actions are morally wrong, but nevertheless exploits others without bad conscience. Or someone suffering from chronic depression, who cannot bring himself to be motivated to act whatsoever, not least in accordance with his moral judgements. If amoralists are possible, which they seem to be, motivational internalism looks implausible.

Many motivational internalists have responded by denying that amoralists are making genuine moral judgements (at least given that their psychology is otherwise normal). See for instance Hare (1963), Smith (1994) and Björklund et al. (2012) for further examples. It remains controversial whether that denial is plausible. Therefore, I briefly want to indicate a different response which avoids these issues. It is possible to account for amoralists while retaining *both* motivational internalism, and the fact that amoralists are making genuine moral judgements while being otherwise psychologically normal. On the present version of motivational internalism, making implicit moral judgments must include having an emotional attitude. But making explicit moral judgements must not include having an emotional attitude. Hence, we could say that the reason amoralists remain emotionally unaffected by their moral judgements is because they are making explicit moral judgements without also making the corresponding implicit moral judgements.

That being said, someone who is still skeptical towards this kind of internalism could deny it and instead account for the explanatory link between one's recognizing the relevant moral facts, one's subsequent concern for the relevant right-making features, and one's subsequent motivation to perform a particular action that is morally right in some other way. Exactly how the explanatory link is to be accounted for remains an open question. But however conditions (1), (2), and (3) of **The Explanation View** are fulfilled, the action in question will end up having moral worth.

Finally, I have assumed that explicit recognition of the relevant moral facts is not necessary for moral worth. But this claim can be disputed. If we compare two agents who perform the same morally right action, where one has explicitly recognized the relevant moral

facts, and the other only has an implicit recognition, should not the agent with explicit recognition be deemed as *more* praiseworthy than someone with mere implicit recognition? In a related discussion of what it takes to act for moral reasons, and its significance for moral worth, Hills (2018) makes a similar point. Hills argues that performing morally worthy actions is not just ‘doing the right thing for the right reasons’, because one can act for the right (moral) reasons without explicitly recognizing those reasons as *moral* reasons. And according to Hills, explicitly recognizing the relevant reasons as *moral* reasons is important for moral worth. In her view, this is because:

...morally ideal action includes responsiveness to morality in as many ways as possible. Morally worthy action is a moral ideal, because it is action which flows from the agent’s orientation towards the good (Hills 2018, p. 261).

Similarly:

...it is possible to do the right thing for the right reasons without being aware that you are doing it. But the most morally admirable action involves a full orientation towards the good, which requires full responsiveness to moral reasons, that is a recognition of them as reasons, including acknowledging their normative force (that they require, permit, or recommend action) and putting that recognition into action. Morally ideal action involves full responsiveness to moral reasons in all relevant ways. Morally worthy action is not equivalent to – because it is more demanding than – doing the right thing for the right reasons (Hills 2018, p. 262).

Translating to present terminology, this is essentially to say that morally worthy action requires explicit recognition of the relevant moral facts. But this assumption is dubious. There are two strong related reasons why explicit recognition is not necessary for moral worth.

First, agents often cannot be expected to have acquired the relevant knowledge or conceptual framework required to have an explicit recognition of the relevant moral facts. In those cases, it is implausible that the praiseworthiness of their actions diminish.

Consider again the Huckleberry Finn-case. Finn does not explicitly recognize the normative force of the fact that Jim is his friend. Still, Finn implicitly recognizes the normative force of the fact that Jim is his friend and acts rightly by helping him because of it.

The praiseworthiness of Finn's helping Jim does not diminish just because Finn's upbringing has not given him access to the relevant knowledge and conceptual framework required to understand this fact explicitly. The initial intuition that most have, namely that Finn's helping Jim has moral worth, further supports this verdict.

Second, if explicit recognition of the relevant moral facts was necessary for moral worth, it would create an implausible elitism about morally worthy actions. Only people who explicitly study morality would be able to perform morally worthy actions, since those are the only ones with access to the relevant knowledge and conceptual framework required to recognize the relevant moral facts explicitly. But this cannot be right. People often perform actions of utmost praiseworthiness, even though they are not moral philosophers nitpicking about whether or not some particular action has the property of being morally right. My father makes me a sandwich without me asking for it, my mother listens to my problems, and my brother takes interest in my work. Neither perform these actions with any explicit recognition of the relevant moral facts. But I cannot see how their actions could be any more praiseworthy than they already are.

## 9. Final thoughts

I have shown that the prospects for the Moral Concept View are highly limited. The Moral Concept View cannot account for the fact that an agent's action can have moral worth even if he does not act from a moral concern. Neither can it account for the fact that an agent's action cannot have moral worth if he *only* acts from a moral concern, without concern for anything else.

I also corrected mistaken assumptions about what it takes to act rightly non-accidentally. In doing so I demonstrated that while the Strong Non-Moral Concept View is untenable, it is indeed possible to explain how agents can non-accidentally act rightly in a way that is compatible with the Weak Non-Moral Concept View. By conceding that **The Strong Claim** is false in favor of **The Weak Claim** and focusing on **Explained**, adherents of the Non-Moral Concept View will be all set to tell a story about the nature of morally worthy actions. However, more work must be done to fully establish **Explained** as not only a sufficient, but also necessary condition for non-accidentally acting rightly. This is a difficult task since these matters are highly controversial. Generating different sufficient conditions for non-accidentally acting rightly might be the best that we can do.

A final note for the future is that any account of moral worth must account for two important features of the agent's disposition: a motivational component and a belief-component. Accounting for the motivational component ensures that the agent's heart is in the right place (so to speak), while accounting for the belief-component ensures that the agent is sufficiently morally aware. I conclude that with respect to the motivational component, the Non-Moral Concept View has the advantage. Caring directly about the actual things of moral importance remains important for moral worth. With respect to the belief-component, the Moral Concept View has the advantage (with some qualifications). Although moral beliefs remain important for moral worth, they need not be explicit.



## References

- Arpaly, N. 2002, "Moral Worth", *The Journal of Philosophy*, vol. 99, no. 5, pp. 223-245.
- Arpaly, N. & Schroeder, T. 2013, "Credit and Blame", in *In Praise of Desire*. Oxford: Oxford University Press.
- Björklund, F. et al. 2012, "Recent Work on Motivational Internalism", *Analysis (Oxford)*, vol. 72, no. 1, pp. 124-137.
- Gettier, E.L. 1963, "Is Justified True Belief Knowledge?", *Analysis (Oxford)*, vol. 23, no. 6, pp. 121-123.
- Hare, R. 1963, *Freedom and Reason*. Oxford: Oxford University Press.
- Hills, A. 2018, "What Does it Take to Act for Moral Reasons?" in *The Many Moral Rationalisms*. Oxford: Oxford University Press.
- Howard, N.R. 2021, "One Desire Too Many", *Philosophy and Phenomenological Research*, vol. 102, no. 2, pp. 302-317.
- Johnson King, Z.A. 2020a, "Accidentally Doing the Right Thing", *Philosophy and Phenomenological Research*, vol. 100, no. 1, pp. 186-206.
- Johnson King, Z.A. 2020b, "Praiseworthy Motivations", *Noûs*, vol. 54, no. 2, pp. 408-430.
- Kant, I. & Wood, A.W. 1785 [2018], *Groundwork for the Metaphysics of Morals: With an Updated Translation, Introduction, and Notes*. New Haven: Yale University Press.
- Kolodny, N. & Macfarlane, J. 2010, "Ifs and Oughts", *The Journal of Philosophy*, vol. 107, no. 3, pp. 115-143.
- Lackey, J. 2006, "Pritchard's Epistemic Luck", *The Philosophical Quarterly*, vol. 56, no. 223, pp. 284-289.
- Lillehammer, H. 1997, "Smith on Moral Fetishism", *Analysis (Oxford)*, vol. 57, no. 3, pp. 187-195.
- Moore, G.E. & Shaw, W.H. 1912 [2005], *Ethics*. Oxford: Oxford University Press.
- Nozick, R. 1981, *Philosophical Explanations*. Cambridge: Harvard University Press.
- Markovits, J. 2010, "Acting for the Right Reasons", *The Philosophical Review*, vol. 119, no. 2, pp. 201-242.
- Olson, J. 2002, "Are Desires De Dicto Fetishistic", *Inquiry (Oslo)*, vol. 45, no. 1, pp. 89-96.
- Pritchard, D. 2005, *Epistemic Luck*. New York: Oxford University Press.
- Riggs, W.D. 2014, "Luck, Knowledge, and "Mere" Coincidence", *Metaphilosophy*, vol. 45, no. 4-5, pp. 627-639.

Schroeder, T. "Desire", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2020/entries/desire/>>.

Sliwa, P. 2016, "Moral Worth and Moral Knowledge", *Philosophy and Phenomenological Research*, vol. 93, no. 2, pp. 393-418.

Smith, M. 1994, *The Moral Problem*. Oxford: Blackwell Publishing.

Sosa, E. 1999, "How to Defeat Opposition to Moore", *Philosophical Perspectives*, vol 13, pp. 141-153.

Star, D. 2008, "Moral Knowledge, Epistemic Externalism, and Intuitionism, *Ratio (Oxford)*, vol. 21, no. 3, pp. 329-343.

Star, D. (ed.) 2018, *The Oxford Handbook of Reasons and Normativity*. New York: Oxford University Press.

Tomlinson, L. 2020, "The Moral Worth of Intentional Actions", *Philosophy and Phenomenological Research*, vol. 101, no. 3, pp. 704-723.

Twain, M. & Worthington, A. 2015, *The Adventures of Huckleberry Finn*. New York: Open Road Integrated Media.

Williams, B. 1981, "Persons, Character, and Morality", in *Moral Luck*. Cambridge: Cambridge University Press.

Williamson, T. 2000, *Knowledge and its Limits*. New York: Oxford University Press.

Zhang, X. 2021, "Why De Dicto Desires are Fetishistic", *Ratio (Oxford)*, vol. 34, no. 4, pp. 303-311.

Zimmerman, M.J. 2002, "Taking Luck Seriously", *The Journal of Philosophy*, vol. 99, no. 11, pp. 553-576.