

# Moral Worth and Moral Knowledge<sup>1</sup>

PAULINA SLIWA

*University of Cambridge*

To have moral worth an action not only needs to conform to the correct normative theory (whatever it is); it also needs to be motivated in the right way. I argue that morally worthy actions are motivated by the rightness of the action; they are motivated by an agent's concern for doing what's right and her knowledge that her action is morally right. Call this the Rightness Condition. On the Rightness Condition moral motivation involves both a conative and a cognitive element—in particular, it involves moral knowledge. I argue that the Rightness Condition is both necessary and sufficient for moral worth. I also argue that the Rightness Condition gives us an attractive account of actions performed under imperfect epistemic circumstances: by agents who rely on moral testimony or by those who, like Huckleberry Finn, have false moral convictions.

## 0. Introduction

Not all right actions are morally praiseworthy. We're hesitant to praise a political candidate who advocates deep cuts to poverty-relief programs for washing pots in a soup kitchen on the campaign trail. There's no question that volunteering in soup kitchens is morally right—perhaps even morally required. But the candidate's political commitments raise doubts about the nature of her motivation.

Whether an action is morally praiseworthy depends not just on whether it conforms to the correct normative theory (whatever it is). It needs to be motivated in the right way. An account of moral worth aims to identify what such good motivations consist in. My aim in this paper is to develop

---

<sup>1</sup> Many people have helped me think through the issues in this paper. I would like to thank Eric Brown, Alex Byrne, Tom Dougherty, Thomas Douglas, Elizabeth Harman, Sally Haslanger, Brian Hedden, Richard Holton, Sophie Horowitz, Rae Langton, Kate Manne, Julia Markovits, Miriam Schoenfield, Thomas Simpson, Michael Smith, Katia Vavova, and Kenny Walden for helpful comments, questions, and conversations. In addition, this paper has benefited enormously from helpful and extensive feedback by several anonymous reviewers. I'm also grateful to audiences at MIT, Oxford, LSE, SLACRR, and Cambridge.

and defend one particular answer to this question. On this answer, morally worthy actions are those that are motivated by the rightness of the action. That is, they are motivated by an agent's concern for doing what's right and her knowledge that her action is morally right. I defend the following:

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).

Call this the Rightness Condition.

I argue that the Rightness Condition is a necessary condition for moral worth. My argument appeals to a central feature of morally worthy actions that has been suggested by Kant:

In the case of what is to be morally good it is not enough that it [i.e. the action] conform with the moral law but it must also be done for the sake of the law; without this that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also often produce actions contrary to law.<sup>2</sup>

The thought is that morally worthy actions are motivated in a way that makes their rightness neither “contingent” nor “precarious”—they are counterfactually robust. I argue that, unlike alternative accounts of moral worth that have been defended in the literature, the Rightness Condition gives us a plausible and principled account of this counterfactual robustness. I defend the Rightness Conditions against objections put forward by, amongst others, Nomy Arpaly and Julia Markovits.<sup>3</sup> Despite requiring knowledge of what the right thing to do is as part of the agent's motivation, the Rightness Condition does not put morally worthy actions beyond the reach of epistemically limited agents like us.

I then argue that the Rightness Condition is sufficient for moral worth by appealing to cases of moral testimony. Moral testimony can be a source of moral knowledge, and so, on the Rightness Condition, it can give rise to morally worthy actions. I argue that while we should welcome this implication, alternative accounts of moral worth—particularly those on which moral worth requires responsiveness to right-making reasons—struggle to accommodate it.

The Rightness Condition tells us when actions have moral worth. But it also has implications for when agents are morally good. I draw out these implications by considering the much-discussed case of Huckleberry Finn.

---

<sup>2</sup> Kant, 3–4 (4:390).

<sup>3</sup> In particular, in Arpaly (2003) and Markovits (2010).

## 1. Some Clarifications

Before we start, two clarifications will be helpful. The first concerns what exactly moral worth is about. I am interested in when it's appropriate to praise an agent in the sense that she is praiseworthy for the action. Praising agents is itself something we do and we may do it for all kinds of reasons other than that the action merits praise. If you hold a gun to my head, I'm going to praise your "exquisite" singing even if your sense of key is questionable. In this case it may be prudentially appropriate that I should praise it. But you are not praiseworthy for your performance: you do not deserve credit for your singing. So, the sense of being praiseworthy that I'm interested in is what we mean when we say that an agent "deserves credit for her success". Even if our actual practice of praising agents is somewhat messy, I believe we have a good enough grip on this notion of praiseworthiness to provide us with a starting point for a philosophical investigation.

Secondly, when I refer to right actions as being morally praiseworthy, I use the term 'morally right' broadly to include not just actions that are morally required but also those that may not be required but are nevertheless morally good. Thus, if an agent has a choice between several equally morally good actions, such as donating to a flood-relief charity or donating to cancer-research, then she does the right thing in donating to cancer-research, even though she is not required to do so—it would have been permissible for her to donate to flood-relief instead. Similarly, if there are supererogatory actions, they are morally right, even if they are not morally required.

## 2. Motivating the Rightness Condition as a Necessary Condition

The Rightness Condition has two distinctive features: first, it insists that for an action to have moral worth the agent must be motivated by mental states with moral content: such as, for example, a desire to do the right thing. Second, the Rightness Condition departs from alternative proposals in what kind of mental states must figure in the agent's motivation in order for her actions to have moral worth. It requires that knowledge—knowledge of what the right thing to do is—must be part of what motivates the agent to act.<sup>4</sup> My aim in this section is to show that this allows the Rightness Condition to capture that the kind of motivation that gives rise to morally worthy actions is one which

---

<sup>4</sup> The Rightness Condition thus crucially relies on the assumption that knowledge is a mental state that can play a causal role in producing actions. While this claim is not uncontroversial, I cannot offer a defense of this paper here. For a defense of this position, see Williamson (2000). See also Gibbons (2001) for an argument that knowledge must play a causal role in producing intentional actions. And see Nagel (2013) for an argument that knowledge is standardly taken to be a mental state in the psychology literature, as well as a discussion of why it's controversial in philosophy.

makes their “conformity with the moral law” non-accidental. As Barbara Herman writes:

...when we say that an action has moral worth, we mean to indicate (at the very least) that the agent acted dutifully from an interest in the rightness of his action: an interest that therefore makes its being a right action the nonaccidental effect of the agent’s concern.<sup>5</sup>

We can get a first impression of this by considering some cases. Take Kant’s shopkeeper who deals honestly with his customers because he wants to keep his business profitable; he doesn’t care that it is the right thing to do. Insofar as the shopkeeper is motivated solely by a desire for profit, his doing the right thing is a consequence of a fortuitous alignment of what suits his desire for profit and what’s morally right. Had it been profitable to cheat his customers, he may have done that instead.

The Rightness Condition rightly excludes such actions from having moral worth since such actions fail to satisfy the conative requirement: a right action has moral worth only if it was motivated by moral concern.

Moral concern is taken as concern for doing what’s right. This is meant to be understood broadly enough to include such conative states as a desire to do what’s right, an intention to do what’s right, etc. What’s important is that the concern for doing what’s right be non-instrumental: the agent must care about doing what’s right for its own sake, and not because it would further some other goal. Thus, the shopkeeper’s honest dealings with a customer are not praiseworthy because insofar as he is motivated by a desire to do what’s right, this desire is merely instrumental.

One kind of accidentality that’s incompatible with moral praiseworthiness is when an agent performs a right action but from an ulterior or selfish desire. But sometimes the conative requirement is met and yet the agent’s right action still seems worryingly accidental. Consider, for example, a bureaucrat like Eichmann, who seems to care about doing what’s right but has a deeply misguided conception of what morality requires: he believes that doing what’s right requires unquestioning obedience in carrying out the orders of his superior.<sup>6</sup> Suppose that, for once, this superior does order our bureaucrat to do something that’s morally right: send some supplies to the needy, for example. Thus, the agent does what’s right and he is motivated by concern for doing what’s right. Yet, intuitively, he is not morally praiseworthy for his action.

---

<sup>5</sup> Herman (1981), p. 366.

<sup>6</sup> Arpaly (2003) discusses a similar case but uses it to argue for a different conclusion: that moral worth requires being responsive to right-making reasons *de re*.

The Rightness Condition rightly excludes such actions from moral worth, too.<sup>7</sup> This is because it includes the knowledge requirement. Depending on how we fill in the details, the bureaucrat may well have a justified true belief about what the right thing to do is. But it's implausible that the bureaucrat could know what the right thing to do is. His superior is not a reliable source of moral guidance. And given the bureaucrat's misguided conception of what is right, he is not in an position to identify reliable advisors on moral questions. Since the bureaucrat does not know what the right thing to do is, it's a matter of luck that he does the right thing. If an agent lacks knowledge of what the right thing to do is, then her successfully doing the right thing is worryingly accidental, even if she wanted to do what's right.

### 3. Why Reason-Responsiveness Will Not Do

The Rightness Condition is controversial. Many recent views of moral worth argue that to perform morally worthy actions an agent does not need to be motivated by any mental states with moral content at all. Rather, she should be directly motivated by concern for those nonmoral features which make her action right—whether or not she knows them to be right-making features. Nomy Arpaly has argued that a necessary and sufficient condition for moral worth is that the agent be motivated by moral concern, understood as follows:

Moral concern is to be understood as concern for what is in fact morally relevant and not as concern for what the agent takes to be morally relevant. [...] To say that a person acts out of moral concern is to say that a person acts out of an intrinsic (noninstrumental) desire to follow (that which in fact is) morality, or a noninstrumental desire to take the course of action that has those features that make actions morally right.<sup>8</sup>

Thus, an agent can be morally praiseworthy as long as she is motivated by a desire to relieve suffering, feed the hungry, help the needy—even if she does not conceive of them as the right thing to do. Similarly, Markovits argues:

According to what I will call the Coincident Reasons Thesis, my action is morally worthy if and only if my motivating reasons for acting coincide with the reasons morally justifying the action—that is, if and only if I perform the action I morally ought to perform, for the (normative) reasons why it morally ought to be performed. My motivating reason for perform-

---

<sup>7</sup> Unlike, as noted by Arpaly (2003), p. 73, Barbara Herman's (1993) account, on which moral worth is a matter of acting in light of one's judgment of what's right.

<sup>8</sup> Arpaly (2003), p. 84.

ing some action in this case will not be the duty-based reason “that the moral law requires it” but the reasons for which the moral law requires it.<sup>9</sup>

My aim in this section is to argue that views on which being motivated by a desire for those features of the situation that make the action right suffices for moral worth—regardless of whether the agent conceives of those features as right-making features—face a serious problem. This problem takes the form of a dilemma. Such views either cannot accommodate the central feature of morally praiseworthy actions: that they are motivated in a way that makes them non-accidentally right. Or they have to give up on the thought that whether an agent is morally praiseworthy for an action only depends on the mental states that actually moved her to act, not on those that would have moved her to act in various counterfactual scenarios. Both options, I will argue, are unattractive.

Consider first an agent who does the right thing and is motivated by an individual *de re* desire for the relevant right-making reason. Jean’s friend missed her bus to work and frets over being late to an important meeting; coming late would be a great embarrassment to her. Wanting to spare her friend a major embarrassment, Jean gives her a ride. Let’s assume that giving her friend the ride is the right thing to do in these circumstances and the fact that it spares her friend a major embarrassment makes it right. Thus, Jean is acting from a *de re* desire for a right-making reason. Does Jean’s action have moral worth?

A central feature of morally worthy actions is that they are not merely accidentally right. Given Jean’s motivation, it’s not a fluke that Jean spared her friend a major embarrassment. But it is a fluke that she did the right thing. This is because there are plenty of circumstances in which the considerations that an action will spare one’s friend a major embarrassment is outweighed by other morally relevant factors. In these cases, Jean’s motivating *de re* desire would lead her to do the wrong thing. For example, Jean ought not murder her friend’s ex boyfriend, even if doing so would eliminate a major source of embarrassment in her friend’s life. Hence, if what motivates Jean is solely a *de re* desire for the particular right-making reason that makes giving her friend a ride right on this occasion, then it’s a matter of luck that she acted rightly. And so, her doing the right thing seems too “precarious” for her action to have moral worth. Thus, if we want to hold on to the thought that morally worthy actions are not just accidentally right, an isolated *de re* desire will not do.

---

<sup>9</sup> Markovits (2010), p. 205. Similarly, Stratton-Lake (2000) argues:

[We] must abandon the idea that a morally good person would be disposed to be motivated to do what she should by the thought of duty, and a fortiori the idea that she would be disposed to be motivated solely by this thought (p. 27).

In response, one might suggest that whether Jean's action has moral worth is not just a matter of the actual desire that motivated her to act. Rather, we need to take into account how Jean would have been motivated in various counterfactual scenarios: we need to look at her motivational set as a whole. Suppose that Jean has a *de re* desire for each of the relevant right-making reasons that might arise in various possible situations (such as when giving a ride would help her friend get a promotion, rather than just save her an embarrassment), plus desires to avoid the possible wrong-making features that might arise (such as a desire to not do any killing on behalf of her friend). Moreover, all these desires have just the right strength, so as to ensure that they lead Jean in various possible situations to act in accordance with what is in fact right. If Jean's psychology were so constituted, then it would be true that it's not an accident that Jean performs the right action; her doing the right thing would be counterfactually robust. Thus we might suggest that whether Jean's action has moral worth not only depends on those mental states that actually moved Jean to help her friend—her desire to save her friend an embarrassment and her belief that giving her a ride will accomplish that. It also depends on the content and strength of other desires that Jean has—desires that would have motivated her to act, had circumstances been different.<sup>10</sup>

A view along these lines faces two difficulties. For one, it just does not seem right that when deciding whether or not to give credit to an agent for an action, we care about anything but the agent's actual motivation. Markovits describes the case of a dog-lover, who risks his life to save a drowning stranger but who could have easily been motivated differently. In particular, had his dog been present, he would have been unwilling to abandon the dog—or perhaps too distracted to notice the stranger. If we insist that what matters for moral worth is not just how an agent was in fact motivated but how she would have been motivated in different circumstances, then we must conclude that the dog-lover does not deserve praise for his heroic deed or that he only deserves minimal praise.<sup>11</sup> But neither of these seem plausible.

In general, when deciding whether to give an agent credit for an action—including nonmoral credit—we are interested in the motivations that

---

<sup>10</sup> Variants of such a view have been defended by Stratton-Lake (2000) as well as by Arpaly (2003). Both accounts take a second-order dispositional motive to be necessary for moral worth. On Arpaly's (2003) account it's important that the agent wasn't just actually motivated by the right-making reasons, but she would have been motivated by right-making reasons even if circumstances had been different. The strength of this second-order dispositional motive determines, on Arpaly's account, the degree of moral praiseworthiness for an action. According to Stratton-Lake (2000), an agent must not only have been motivated by right-making reasons, but it must be true that she wouldn't have performed the action if she had judged it to be wrong.

<sup>11</sup> Markovits (2010), p. 210.

in fact led the agent to act. Consider a chess player who responded to her opponent's move in the right way, thereby saving her queen: she wanted to win and hence to save her queen and she knew how to do it. If that's what motivated her move, we give her credit for it; it's hardly a fluke that she succeeded in responding correctly to the threat at hand. We do not give her any less credit for it just because, had she suffered from a bout of insomnia before the game, she would have been motivated differently and played badly as a consequence: either because she would have been too tired to care about winning the game or because she would have been too fuzzy-headed to think of the correct move.

There is a second serious challenge for views on which an action's moral worth is not only a function of the agent's actual motivation but also of how the agent would have been motivated in various counterfactual scenarios: clearly, it is unreasonable to demand that to have moral worth the agent needs to have acted rightly no matter what. Some contingency must be compatible with moral praiseworthiness. The question is: where do we draw the line? Markovits rightly notes:

We all have our breaking points, whether they're triggered by threats to our own interests or to the interests of those we love. So a criterion for moral worth according to which our being motivated by the right-making reasons would have to be completely independent of contingent circumstances for our acts to count as morally worthy entails that virtually no acts at all would qualify.<sup>12</sup>

Markovits concludes that since there is no principled line to be drawn, we should give up on counterfactual robustness as a mark of moral worth altogether.<sup>13</sup> I think Markovits is right that views she targets—in particular Arpaly's and Stratton-Lake's accounts—do not succeed in drawing such a principled distinction between those counterfactuals that matter and those that do not. But abandoning the thought that morally worthy actions are non-accidentally right completely strikes me as too high a price to pay.

#### **4. The Rightness Condition and Counterfactual Robustness**

Fortunately, it's a price we don't have to pay. Unlike its competitors, the Rightness Condition ensures that there is a counterfactually stable link between an agent's actual motivation and her right action. This is because, on the Rightness Condition, morally worthy actions are motivated by both a desire to do what's right and knowledge of what the right thing to do is. But knowledge is by its nature both factive and counterfactually robust: if

---

<sup>12</sup> Markovits (2010), p. 212.

<sup>13</sup> Markovits (2010), p. 210.



an agent knows that an action is the right thing to do, then it is the right thing to do and she could not have easily been mistaken about it's being the right thing to do.<sup>14</sup> And so, if an agent is motivated by concern for doing what's right and knowledge of what the right thing to do is, then it's not a fluke that she acts rightly.

Thus, suppose that Jean gave her friend a ride because she wanted to do what's right and she knew what the right thing to do was: namely, to give her friend a lift. Given this motivation, it's not an accident that Jean did the right thing; there is no possible world in which this motivation would give rise to a wrong action. Contrast this with a case in which Jean is motivated by the desire to save her friend an embarrassment, along with a belief that giving her a ride is a way of doing so. In this case there are a number of possibilities in which this very motivation will give rise to a wrong action—namely, all those scenarios in which the reasons for saving her friend an embarrassment will be outweighed by other moral considerations.

The Rightness Condition also gives us a principled way of pinning down the nature of the counterfactual robustness of morally worthy actions. Morally worthy actions inherit their counterfactual robustness from the knowledge requirement; they are non-accidentally right because they are motivated by moral knowledge and concern for doing what's right. They do not have moral worth in virtue of being counterfactually robust in certain ways. Unlike accounts of moral worth that require a second-order dispositional motive, the Rightness Condition does not have to stipulate an arbitrary cut-off for just how much counterfactual robustness is required for an action to have moral worth. The counterfactuals that matter are simply those that come from our best account of knowledge. And the Rightness Condition rightly excludes odd and deviant cases, in which the agent couldn't have done but the right thing—for example, in which the agent correctly guessed what the right thing to do is but a scheming demon would have intervened if she hadn't guessed correctly.

Let's review where we are. So far, I have argued that the Rightness Condition is a necessary condition for moral worth: an agent deserves moral credit for her right action only if she is motivated both by moral concern—a conative attitude with moral content—and by moral knowledge. On the Rightness Condition then, morally worthy actions are counterfactually robust along two independent dimensions: the agent's desire to do what's

---

<sup>14</sup> The question how exactly to characterize the counterfactual robustness of knowledge is subject to on-going debate. But there is wide agreement that knowledge implies some counterfactual stability. Williamson (2000), in particular, argues for safety as a condition for knowledge. See also Hawley (2003) for a discussion of counterfactual stability of knowledge-how.

right means that she would not have performed the action if she had believed it to be wrong. The agent's moral knowledge means that she would not have been easily mistaken about what the right thing to do is. And so, there is a counterfactually robust link between the agent's motivation and her doing the right thing. In what follows, I will defend the Rightness Condition against some objections. I tackle objections to the knowledge requirement first, before discussing a central objection to the conative requirement.

## 5. Defending the Knowledge Requirement

According to the Rightness Condition, an action has moral worth only if it's motivated by a conative state with moral content as well as knowledge that it is the right thing to do. This is less demanding than rival accounts of moral worth that place constraints not only on how the agent was in fact motivated but also on how she would have been motivated in various counterfactual scenarios. Nevertheless, the knowledge requirement in particular has been criticized for making morally worthy action too hard to achieve. In this section I hope to dispel these worries. I will first argue that the Rightness Condition accommodates the intuition that there is often plenty to admire when agents, through no fault of their own, fall short of knowing what the right thing to do is. Second, I will argue that nothing weaker than knowledge will do: to settle for justification, or justification and truth is to give up on the thought that morally praiseworthy actions are non-accidentally right.

According to the Rightness Condition, an agent is morally praiseworthy for an action only if it's the right thing to do. This may strike you as too demanding. Knowing what the right thing to do is often requires the agent to have nonmoral knowledge about the situation. Agents may sometimes be blamelessly ignorant about their circumstances and act wrongly as a result. And in such cases, you might argue, it's ungenerous to deny them praise. Consider the following example:

Peter reads in a reliable source that giving to charity X is a very effective way of supporting famine-relief: this charity distributes excess US corn amongst the needy in the developing world. Since Peter takes supporting famine-relief to be the right thing to do, he decides to give money to this charity. As it happens, Peter's belief that donating to charity X is the right thing to do is false. This is because distributing excess US corn exacerbates famine problems; it puts local farmers out of business. But Peter does not know that.

Plausibly, it's morally wrong to support an organization that exacerbates famine. But it seems both implausible and ungenerous to deny that there is

something morally admirable about Peter's action. After all, Peter seems to be acting on good reasons; he is justified in believing that supporting charity X is the right thing to do. And so, one might conclude that the Rightness Condition is simply too demanding as a criterion for when agents deserve moral credit for their actions.

This conclusion is too quick. It's true that, according to the Rightness Condition, Peter is not morally praiseworthy for supporting charity X. This is because supporting charity X is morally wrong. However, this doesn't mean that there is nothing morally praiseworthy about what Peter does. After all, Peter performs a number of actions in supporting charity X. By donating money, he gives a part of his income to charity. He also follows reliable advice about how to support famine-relief. And, despite the fact that he does not succeed, he attempts to support famine-relief efforts in the developing world. Giving away part of one's income to charity is the right thing to do and plausibly Peter knows that. In situations, in which you are not in a position to evaluate the evidence yourself, following expert advice about which charity to support is also the right thing to do. And again, plausibly Peter knows that. The same is true for attempting to support famine-relief efforts. Thus, while Peter may have acted wrongly in supporting charity X, he did act rightly in donating part of his income, in following advice about which charity is effective, and in attempting to support famine-relief efforts. Insofar as Peter knows that these actions are right, he may well be morally praiseworthy for them.

These considerations highlight that we must take care in identifying exactly which action agents are morally praiseworthy for. Since most actions are complex, agents who perform some morally wrong action may, at the same time, perform actions that are morally right. When these actions are motivated in the right way, the agent is morally praiseworthy for them. The Rightness Condition can thus accommodate the intuition that there is often something admirable about agents, like Peter, who act wrongly.

According to the Rightness Condition, moral worth requires moral knowledge: knowledge of what the right thing to do is. But couldn't an agent perform morally worthy actions even when she does the right thing despite having false moral beliefs? Suppose, for example, that some form of deontology is the correct moral theory. Does it follow that a consequentialist is precluded from performing actions that have moral worth?

Even if deontology is the correct moral theory, it doesn't follow that a consequentialist cannot have moral knowledge. When you are trying to determine what the right thing to do is, knowing the correct moral theory is, of course, very helpful. But it is not necessary: often even a false moral theory can be a reliable guide to what the right thing to do is. And so, even a false moral theory can be a source of moral knowledge. After all, even a false scientific theory can be a source of a great deal of scientific

knowledge. Plausibly, someone who uses Newtonian mechanics to determine whether a bridge will be stable, and who performs the calculations correctly, can thereby come to know that it will be stable—even though Newtonian mechanics, as a scientific theory, is false. In the same way, a false moral theory can function as a reliable heuristic for what the right thing to do is. Even if there is some disagreement between consequentialists and deontologists, there is also a lot of convergence.<sup>15</sup> More generally, the Rightness Condition allows that an agent can be morally praiseworthy for a right action even when she is ignorant of the correct moral theory. This is because knowing what the right thing to do is in a particular situation is compatible with being ignorant as to what the correct moral theory is.<sup>16</sup>

At the same time, the Rightness Condition avoids a problem that besets accounts on which all that's required for moral worth is acting on a justified belief about what the right thing to do. This has been suggested by Markovits:

Because the reasons relevant to moral-ought claims are subjective—they depend on what an agent ought to believe about her situation—our normative reasons for acting can't be given by facts of which we're blamelessly ignorant [...].<sup>12</sup>

Such an account gives implausible verdicts for agents who do the wrong thing based on a justified but false moral belief. Thus, imagine Ann who grows up in a very remote and tight-knit community. Her friends and neighbors are honest and kind people. Unfortunately, there is consensus in Ann's community that gay marriage is a great moral evil. Ann has never met a gay person and her belief is not based on hatred or dislike of gay people. Rather, she has acquired her belief that same-sex marriage is a moral evil along her other moral beliefs, such as that one should keep one's promises or that one should not be cruel to animals. It strikes her as intuitively obvious that same-sex marriage is "different" from "regular" marriage and that this difference is morally relevant.

It seems plausible that Ann could be blameless for her moral ignorance. Her moral ignorance is the result of epistemic bad luck: her sheltered upbringing. She has no reason to distrust the community consensus or her own intuitions on this particular issue and those around her are morally reliable on

---

<sup>15</sup> Even if we disagree with Parfit (2011), vol. 1 that consequentialists, Kantians, and contractualists are all "climbing the mountain on different sides" (p. 419), it seems plausible that at least a substantial part of the disagreements between different normative theories concern either the justification of why an action is the right or whether it's required or supererogative and not whether it's morally right.

<sup>16</sup> I'll discuss the Huckleberry Finn case in the final section.

other moral questions. And so, it seems, Ann's belief that same-sex marriage is a moral evil, while misguided and unfortunate, may well be justified.

Now suppose that Ann signs a petition against same-sex marriage. By her own lights this is the right thing to do. Thus, if justification suffices for moral worth, it seems that Ann is morally praiseworthy for signing a petition against same-sex marriage. But this cannot be right. Signing a petition against same-sex marriage is morally wrong. Insofar as Ann's belief that she's doing the right thing is sincerely held and justified, it may be plausible that we should not blame her for her moral mistake. Surely, however, we should not give her moral credit for signing the petition.

Could one maintain that only false nonmoral beliefs are compatible with morally worthy actions? This seems worryingly ad hoc; the asymmetry calls for an explanation.<sup>17</sup> And we can point to a principled worry for this line of response: the motivation for preferring justification to knowledge was that agents sometimes fall short of knowing what the right thing to do is through no fault of their own; when agents do the best they can, it would be ungenerous to deny them moral praise. But then it's unclear why we would distinguish between moral and nonmoral ignorance. Ann is no more at fault for growing up in a community with false beliefs about same sex marriage than Peter is at fault for having misleading information about famine-relief. Both are acting in light of what they have reason to believe is the right thing to do.

In contrast, the Rightness Condition explains why Ann fails to be morally praiseworthy for signing a petition against same sex marriage without having to appeal to any asymmetry between moral and nonmoral ignorance: signing the petition is morally wrong. Insofar as her belief that same-sex marriage is morally impermissible is epistemically justified, she may well not be blameworthy for signing the petition. Doing the best you can, given your epistemic situation, may well excuse you from blame. But it's not enough to earn you moral praise when, despite your best effort, you get it wrong.

Anna's case suggests that merely having a justified belief that an action is right cannot be enough for moral worth. But you may still question whether we need to require knowledge of what the right thing to do is. Wouldn't it be enough for moral worth if an agent was motivated by a true justified belief that her action was morally right?

If it's a central feature of morally worthy actions that they are not just right but non-accidentally so, justified true belief in the absence of knowledge will not do. This is because when an agent's justified true belief that an action is right falls short of knowledge, her doing the right thing looks worrisomely accidental.

---

<sup>17</sup> One way to develop such a strategy is to argue that all instances of false pure moral belief are themselves morally blameworthy. This has been suggested by Harman (2011) and (forthcoming).

Consider a doctor, who, based on her patient's lab results, diagnoses her with iron deficiency and prescribes a nutritional supplement, recommending a common brand of it to the patient. Unbeknownst to her doctor, the lab technician made a mistake. The patient is doing fine on her iron level but she suffers from a vitamin B12 deficiency. Fortunately, the particular brand of nutritional supplement includes both iron and vitamin B12. The doctor's belief that the particular supplement will help the patient is thus true. It is also justified—after all, the lab results provide the doctor with good reason for believing that the patient suffers from an iron deficiency. But does the doctor deserve credit for prescribing the correct nutritional supplement? Intuitively, not. While she did get the treatment right, her success was an accident. This is because the doctor did not know that the patient had a B12 deficiency and hence did not know that the nutritional supplement was the correct treatment.<sup>18</sup> This, of course, is compatible with the doctor's being praiseworthy for a number of other actions: for ordering the correct blood test or for following medical guidelines.

In general, when a justified true belief fails to be knowledge this is precisely because the agent's epistemic circumstances are precarious. To settle for a justified true belief about what the right thing to do is, rather than knowledge, is thus to give up on the thought that when acting in a morally praiseworthy way, the agent's motivation is such that it's not just an accident that she did the right thing.

## 6. Defending the Conative Requirement

The Rightness Condition requires that the agent be motivated by conative attitudes with moral content, such as the desire to do the right thing. In this section I want to address one important objection to this requirement: that there is something morally unattractive about agents who are moved by desires with moral content. And since morally worthy actions are supposed to reflect well on the agent, it's implausible that moral worth could require the agent to be motivated by a desire to do what's right. This objection goes back to Michael Smith's charge that agents motivated by a desire to do what's right are guilty of moral fetishism:

Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, [...] not just one thing: doing what they believe to be right, where this is read *de dicto* and

---

<sup>18</sup> My examples and discussion here follow Gibbons (2001). Gibbons uses them to argue for a different conclusion: that knowledge is necessary for intentional action. I'm not committed to this conclusion here; my argument only appeals to the claim that, without knowledge, the action's success is accidental in a way that's incompatible with praiseworthiness. See also Hawthorne and Stanley (2008).

not *de re*. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue.<sup>19</sup>

I want to start by clarifying what exactly it is that the Rightness Condition requires in terms of a conative attitude. Both Smith and Arpaly, in criticizing agents motivated by moral concern, shift between talking about concern for doing the right thing and concern for doing what one believes to be the right thing. Thus, Arpaly objects to Herman, who argues that to have moral worth an agent's motivation must involve "an interest in the rightness of his action" as follows:

Herman is mistaken if "an interest in the rightness of his action" is interpreted in the most obvious way—that is, as an interest in doing the right thing or the moral thing under this description, in a *de dicto* sense: a concern for doing what one feels or believes, even as a background belief, that one morally ought to do.<sup>20</sup>

Smith, too, says that caring to do what one "believe[s] to be right" is morally objectionable. But a desire to do what's right is not a desire to do *what one believes* is right and I suspect that part of the appeal of this objection derives from this conflation. Of course, if Jean wants to do what's right, she will tend to act in accordance with her judgments and beliefs about what morality requires. But if it turns out that her judgment led her astray, Jean's desire will, by her own lights, not have been satisfied. If, on the other hand, Jean desired to do what she judged to be right, she would receive news of having made a mistaken judgment with perfect equanimity: "Who cares if I was mistaken about what is right. I succeeded in performing the action that I judged to be right; this is all I wanted." Clearly, there is something morally unattractive about such an agent. Even if she acted rightly, her actions would hardly strike us as morally admirable. The Rightness Condition explains why: the conative condition requires an agent to act from a desire to do what's right, not from a desire to do what she judges to be right.

Still, isn't there something intuitively unattractive about an agent who cares about doing what's right? This, I think is much less clear and has been subject to much debate in the literature.<sup>21</sup> A full discussion of these arguments would require more space than I have here. And so, I will limit myself

---

<sup>19</sup> Smith (1994), p. 75. Markovits (2010), p. 204 echoes Smith to dismiss the requirement that agents be motivated by mental states with moral content. For discussion of the same point by Arpaly, see her (2003), pp. 67–69. See also Williams (1981).

<sup>20</sup> Arpaly (2003), p. 73.

<sup>21</sup> See, in particular, Svavarsdottir (1999) for a careful and detailed discussion of Smith's argument as well as a defense of the claim that there is nothing wrong with an agent who is motivated by moral concern *de dicto*. See also Lillehammer (1997), Dreier (2000), and Enoch (2011), pp. 255–56.

to pointing to two reasons for why conative states with moral content are an indispensable part of the motivational set of a morally good person.

First, conative states with moral content are essential for doing the right thing in the face of moral uncertainty. When confronted with murky evidence or with a morally complex situation, a morally good person will be motivated to find out what the right thing to do is. This may require her to collect more evidence about the situation, to engage in moral deliberation, or to seek out moral advice. But it's hard to see she could be motivated to find out what the right thing to do is unless she has some conative states with moral content in her motivational set.

Second, plausibly a morally good person not only reliably does the right thing; she also manifests a range of reactive attitudes to her own actions and those of others. There would be something very odd about an agent who, for example, was neither disposed to experience remorse for her own wrong-doing nor moral indignation when confronted with injustice and wrongdoing by others. Such an agent would strike us as cold and morally unattractive. But to be disposed to experience such reactive attitudes, an agent must conceive of actions in moral terms: as right or wrong.

Why do reactive attitudes such as remorse, moral regret, indignation, require the agent to conceptualize her action in moral terms? Consider what is involved in feeling remorse. To feel remorse is not just a matter of feeling bad or unhappy. Nor is it a matter of feeling this way because one's desires have not been satisfied. I may feel frustrated when I didn't succeed in satisfying my desire to buy avocados because the store had run out of them. And I may feel unhappy because my desire to help my friend to get to work on time was frustrated by the massive traffic jam that we got stuck in. But this sense of frustration or unhappiness is not remorse. Remorse requires conceiving of one's actions as contrary to what one should have done, not merely as contrary to what one wanted to do. It involves feeling bad because one considers oneself as failing with respect to a moral standard. This means that to be disposed to feel remorse for failing to perform an action, an agent must conceive of it as something that she morally ought to do and not just as something she wants or feels compelled to do. And she must care about doing it because it's the right thing to do, i.e. under the relevant moral description. Feeling remorse essentially involves both affective, cognitive, and conative components.<sup>22</sup>

---

<sup>22</sup> A number of authors have defended accounts of reactive attitudes on which they involve cognitive states with moral content. For example, Wallace (1996) argues that "episodes of guilt, resentment, and indignation are caused by the belief that an expectation to which one holds a person has been breached" (p. 12). Sher (2006) also offers an account of blame on which blaming someone requires both beliefs and desires with normative content. Fricker (2014) argues that remorse, in particular, involves mental states "bearing the . . . moral content (X wronged Y)".



## 7. Is the Rightness Condition Sufficient?

I have argued that moral concern and moral knowledge are both necessary for moral worth. Does the Rightness Condition also give us a sufficient condition? While worries about whether the Rightness Condition is necessary focus on the question whether it's too demanding, concerns about whether it's sufficient have focussed on the question whether it's demanding enough. Hills has recently suggested that an account like the Rightness Condition would make moral worth too easy to come by. Consider the following case:

Ron is an extremist, believing that killing a person is not generally immoral but that killing a fellow Jew is a grave sin. Ron would like to kill Tamara, but he refrains from doing so because he wants to do the right thing, and he knows (on the basis of his rabbi's testimony) that the right thing to do is to refrain from killing her.<sup>23</sup>

Hills argues that, intuitively, Ron is not morally praiseworthy for resisting his desire to kill Tamara. But, she argues, Ron does care about doing what's right and based on the rabbi's testimony he knows what the right thing to do is. And so, the Rightness Condition cannot give us a sufficient condition for moral worth. The culprit, Hills suggests, is the knowledge requirement. Ron may have moral knowledge but he fails to have moral understanding and it's the latter that matters for moral worth. To have moral understanding, an agent must have the abilities to give and follow explanations why her action is right and to recognize what the right thing to do is in similar circumstances. Moral understanding is hence considerably more demanding than knowing what the right thing to do is.<sup>24</sup>

I do not think that the case gives us reason to abandon the Rightness Condition as a sufficient condition. Hills' case is very unusual and under-specified. Why does the question whether he may kill Tamara even arise for Ron? Why does he think that killing people, as long as they are not fellow Jews, is morally permissible? Circumstances in which a minimally decent moral agent might genuinely wonder whether it's morally permissible to kill another person are rare.

To come to know what the right thing to do is based on testimony, Ron must be in a position to identify a reliable advisor. But if Ron is so morally incompetent that he needs help to decide whether he may kill another

---

<sup>23</sup> Hills (2009), p. 115. A variant of the case appears originally in Arpaly (2003), who initially uses it to defend a weaker claim: that justified true moral belief along with a desire to do is right is not sufficient for moral worth. Arpaly's later discussion (particularly on p. 73) suggests that she endorses the stronger conclusion also: that even desire to do what's right along with moral knowledge are not sufficient for moral worth.

<sup>24</sup> Hills also argues that it's distinct from simply knowing why something is the right thing to do. See her (2009), pp. 103–106.

human, how could he possibly be in a position to identify an advisor who can be a reliable source of moral guidance?

Thus, insofar as Hills is stipulating that Ron acquires moral knowledge based on testimony despite his lack of basic moral competence, she is describing a case that is subtly incoherent. And if she is not stipulating that, then the case is so unusual that it's too under-described to gauge our intuitions. Even if there is a way to fill in the story so that it's plausible that Ron believes that there is nothing morally objectionable with killing a fellow human and yet that he is morally competent enough to identify a reliable moral advisor, it's not at all clear that with the relevant details filled in, it will still yield the intuitive judgment that Ron does not deserve moral credit for his action.

As described Ron's case doesn't impugn the Rightness Condition. On the contrary, it seems quite natural to appeal to Ron's lack of moral knowledge to explain why he isn't morally praiseworthy for doing the right thing. Given Ron's moral incompetence, he was lucky to come across a reliable advisor and to do the right thing. Ron's case then lends further support to the knowledge requirement as a necessary condition for moral worth.

Ron's action lacks moral worth not because he acts on moral testimony but because he fails to have moral knowledge. But sometimes agents do gain knowledge of what the right thing to do is by relying on others. In such cases, the Rightness Condition implies that their right actions can have moral worth. This is a welcome consequence. Consider the following case:

Anna's older sister is struggling with alcohol addiction; she lost her job, blew through her savings and is several months behind on rent. She asks Anna to "loan" her some money. Anna is conflicted. On the one hand, she does not want her sister to end up homeless. But she wonders whether her sister needs to feel the full consequences of her addiction to finally seek treatment. Moreover, Anna's financial circumstances are modest and she has her own family to look after. Anna is uncertain about what the right thing to do is. She turns to a friend whom she knows to be trustworthy and to have good judgment for advice. Her friend tells Anna that she shouldn't give her sister money. Anna's friend is right: the moral considerations against giving her sister money do outweigh those in favor of it. Anna trusts her friend and acts on her advice. Although it's hard on her, she stays firm and resists her sister's rage and pleas for help.<sup>25</sup>

Intuitively, Anna is morally praiseworthy for doing the right thing and refusing to give her sister money. This is so even if Anna seeks out specifically moral advice. Thus, we can imagine that she knows all the non-moral facts that are relevant to making the right decision—she has studied the

---

<sup>25</sup> This case is similar in spirit to one discussed by Jones (1999).

addiction literature and talked to therapists and social workers about the likely outcomes of both giving and refusing to give her sister money. Nevertheless, she may be uncertain about what the right thing to do is because she may be uncertain how to weigh these considerations (her sister's well-being, Anna's obligations to her own family). In addition, Anna may also have good reason to worry that she might be biased. In difficult situations like this, it often makes sense to rely on advice about what we ought to do rather than try to forge ahead on our own.

Anna's case is very different from Ron's. Unlike Ron, Anna's moral uncertainty does not reflect general moral incompetence. Anna has reason to doubt her own moral judgment about this particular case: the situation is complex, it's novel, and it's very personal. If anything, the fact that Anna seeks out moral advice confirms that she has generally good moral judgment: she is clear-sighted about her own moral limitations. And so, Anna may well be in a position to recognize who could give her reliable advice on that particular question—even when she is not in a position to know what the right thing to do is herself.<sup>26</sup>

Anna seems morally praiseworthy for doing the right thing in refusing to give her sister money because it's not just a fluke that Anna gets it right. She is motivated by concern for doing the right thing. And given that Anna's friend is trustworthy and reliable, it's plausible that, by relying on her testimony, Anna comes to know that refusing to give money to her sister is the right thing to do.<sup>27</sup> Since she knows what the right thing to do is and wants to do what's right, it's hardly accidental that she acts rightly. While the Rightness Condition allows that some actions on moral testimony have moral worth, it doesn't follow that all such actions do.<sup>28</sup>

---

<sup>26</sup> This verdict does not require us to take a stand on whether there are genuine moral experts to which one should always defer. This is because Anna's friend need not be an expert in any such strong or global sense for Anna to reasonably defer to her on this particular occasion—she just needs to be trustworthy and reliable when dealing with the kind of situations that Anna faces. See Jones (1999) and Sliwa (2012) for how one could acquire such more local expertise. See Enoch (2014) for a discussion of how we could recognize someone more reliable than us with respect to a moral question.

<sup>27</sup> This assumes that we can acquire moral knowledge on the basis of moral testimony. But can we? For a defense of this claim see Sliwa (2012). But note that while a number of people have argued that moral testimony is problematic (see Hopkins (2007), McGrath (2011), Howell (2014)), most don't think that failure to transmit knowledge is the problem. An exception to this is McGrath (2009).

<sup>28</sup> This has an important upshot for the debate about moral testimony: one strategy to explain why there is something odd about moral testimony is to argue that actions on moral testimony are precluded from having moral worth. Insofar as one shares the intuition that moral testimony is problematic, one will have to look for an alternative explanation of that fact. For skepticism about the intuition, see Sliwa (2012). See Enoch (forthcoming) for an argument that relying on moral testimony is sometimes required.

Cases like Anna's create problems for alternative accounts of moral worth, on which being motivated by right-making reasons *de re* is necessary for moral worth, as defended by Arpaly and Markovits.<sup>29</sup> The problem is that there isn't an account of right-making reasons which will yield the right verdict in cases like Anna's without committing one to unpalatable consequences in other cases.

According to Arpaly, right-making reasons are those features of an action that explain why it is the right thing to do and an agent needs to be motivated by concern for these features.<sup>30</sup> But reliable moral testimony doesn't explain why an action is right. What explains why Anna should not give her sister money are the non-moral facts about Anna's situation—the fact that doing so would make Anna's sister worse-off in the long run, for example. Moral testimony isn't a right-making reason; it's evidence about what right-making reasons there are and what they on balance support. Thus, when Anna does the right thing because she wants to do what's right and knows, based on her friend's testimony, what that is, she is not acting on right-making reasons *de re*. She is acting from a desire with moral content along with moral knowledge: knowledge what the right-making reasons support. And so, on Arpaly's view, her action lacks moral worth.<sup>31</sup> This is an implausible result.

Markovits seeks to accommodate cases like Anna by adopting a more liberal account of right-making reasons. Markovits suggests that a feature of a situation is a right-making reason if it provides evidence for an action's rightness.<sup>32</sup> While this account of right-making reasons allows Markovits to accommodate actions like Anna's, it has troubling consequences elsewhere. We can easily imagine cases in which the fact that an injection is more painful than the alternatives may be evidence for its effectiveness. This is because, for a fact X to be evidence for another fact Y, it's enough that X reliably correlates with Y. And so, if the painfulness of an injection reliably correlates with its effectiveness, then the fact that injecting you with the drug will cause you pain can be a right-making reason: it can make it right to give you the injection. Now imagine a sadistic doctor who prescribes the painful injection because it's painful and because she non-instrumentally cares about causing you pain (she doesn't at all care about

---

<sup>29</sup> Markovits (2010) argues that her Coincident Reasons Thesis gives us both necessary and sufficient conditions for moral worth. Arpaly (2003) offers "Praiseworthiness as Response to Moral Reasons" (p. 84) as a necessary and sufficient condition for moral worth.

<sup>30</sup> Arpaly (2003), p. 84.

<sup>31</sup> See Arpaly (2003), p. 73.

<sup>32</sup> Markovits (2010), see particularly her discussion on p. 219. Markovits' account of right-making reasons is close to the one defended in Kearns & Star (2009).

the injection being effective). On Markovits' account, such an agent is motivated by non-instrumental concern for a right-making reason. Thus, according to Markovits, the sadistic doctor's prescribing you the painful injection has moral worth. This strikes me as an unacceptable consequence.

The Rightness Condition avoids these unpalatable consequences. At the same time it does not deny that being responsive to right-making reasons is generally important for being in a position to perform morally praiseworthy actions. Morally relevant features—features that either explain the rightness of an action or that reliably correlate with an action's being right—often comprise one's moral evidence. Being responsive to our moral evidence is important for moral knowledge.

I have argued that the Rightness Condition gives us a sufficient condition for morally worthy actions. This is because if the agent is motivated by concern for doing what's right and knowledge of what morality requires, then there is a counterfactually stable link between the agent's right action and her motivation. The Rightness Condition gives us a principled account of this counterfactual link. It also allows that agents in many different kinds of epistemic situations can act in morally worthy way; moral knowledge can come from many sources—deliberation, testimony, or first-hand evidence.

## **8. Being Good and Acting Well**

We can evaluate whether actions have moral worth. But we can also morally evaluate agents: we can evaluate how morally good someone is. While accounts of moral worth tell us how to do the former, they plausibly have implications for the latter. The aim of this section is to outline these implications. At the same time, it responds to a central objection to the Rightness Condition, the much-discussed case of Huckleberry Finn.

There is an important difference between what matters for the moral evaluation of an action and what matters for the moral evaluation of an agent. Whether a particular action has moral worth depends only on whether it was motivated in the right way: by moral concern and knowledge that it's right. Even deeply flawed agents can perform an action that has moral worth when they are moved by a desire to do what's right along with knowledge of what the right thing to do is. In contrast, whether and the extent to which an agent is morally good depends on the pattern of her motivation. Being good is something that admits of degrees. Agents can be more or less morally competent; they can know what's right in some situations while being prone to moral mistakes in others. They can care more or less deeply about doing what's right.

The Rightness Condition suggests that the goodness of agents is a function of their concern for doing what's right and their moral competence:

An agent is morally good to the extent to which she both cares about doing what's right and she is morally competent, i.e. she is in a position to know what the right thing to do is.

On the Rightness Condition, moral praiseworthiness of actions does not admit of degrees. Rather, how virtuous an agent is affects which morally praiseworthy actions they are in a position to perform. A morally good person is in a position to know what's right even in tricky situations, in which those less virtuous are at a loss. This is why morally good people make good moral advisors. Second, their depth of moral concern enables them to perform right actions that would be difficult for others: such as risking one's life to save a stranger.<sup>33</sup> The thought that the moral goodness of an agent could depend on the strength of an agent's moral concern and her moral competence has recently come under attack. It has been argued that it yields implausible verdicts for the case of Huckleberry Finn. A quick reminder about the case:

Huckleberry Finn escapes his abusive father. He meets Jim, a fugitive slave and together they embark on a trip down the Mississippi river on a raft. They make it through quite a few adventures together. Then, suddenly, it occurs to Huckleberry that helping a fugitive slave is like "stealing". In the grip of his ill-trained conscience, Huckleberry resolves to turn Jim over to the authorities. But when the crucial moment comes, Huckleberry finds that he cannot go through with his resolution. Instead, he makes up an elaborate story that protects Jim. In doing so, he clearly does the right thing. However, he believes that he is acting wrongly: he continues to believe that he is complicit in "stealing property" from Jim's "rightful owner", Miss Watson.<sup>34</sup>

Arpaly argues that Huckleberry presents the Rightness Condition with the following challenge: if the Rightness Condition is correct, then moral ignorance is incompatible with morally worthy actions. If we take moral ignorance to be incompatible with morally worthy action, then we have to regard it as incompatible with being a morally good person. And if we regard moral ignorance as incompatible with being a good person, then Huckleberry Finn turns out to be a bad person. But this is implausible:

---

<sup>33</sup> The Rightness Condition thus departs from the "battle citation" model of moral worth, discussed by Henson (1979), on which attributing moral worth to actions "acknowledges a moral victory against odds" (p. 50). The odds need not have been unfavorable for an agent to deserve moral credit for her right action. But of course there may be good pedagogical reasons for giving the agent more praise in such circumstances.

<sup>34</sup> The episode appears in Twain (1884), Chapter XVI. In the context of the moral worth debate, the case is discussed first in Bennett (1974).

Huckleberry Finn [...] is not a bad boy who has accidentally done something good, but a good boy.<sup>35</sup>

Arpaly argues that this problem generalizes: many people hold profoundly misguided moral views and yet strike us as fundamentally decent and good.<sup>36</sup>

On the Rightness Condition, whether Huckleberry's action has moral worth depends on how it was motivated. Insofar as Huckleberry does not know that helping a fugitive slave is the right thing to do—insofar as he saves Jim believing that he is acting wrongly—he is not morally praiseworthy for helping a fugitive slave.<sup>37</sup> But to determine whether Huckleberry Finn is a “good boy”, it's not enough to look at whether he acted in a morally praiseworthy way on a particular occasion. Rather, we need to evaluate the depth of his moral concern and his moral competence. We need to look at the pattern of his motivations; and so, we need to turn to the original story.

As a matter of fact, we find that throughout the story we are given plenty of evidence that Huckleberry is morally competent—his false beliefs about slavery none withstanding. Huckleberry generally knows to treat others fairly, he knows to keep his promises, he knows when to be loyal. We are also given evidence that Huckleberry cares about doing what's right: after all, he apologizes after treating Jim badly. But a genuine apology requires both moral insight and moral concern—it involves recognizing that one has acted as one should not have and that for this reason one ought to make amends.<sup>38</sup> Agents can be generally morally competent even when, for example, because of a racist upbringing, there is a class of situations in which they are systematically mistaken about what the right thing to do is. This systematic moral ignorance is a moral flaw: after all, it led Huckleberry to his morally misguided decision to turn Jim in. But being fundamentally good is compatible with falling short of moral perfection. The Rightness

---

<sup>35</sup> Arpaly (2003), p. 78. Driver (2001), p. xvi, objects to regarding moral knowledge as central to virtue on similar grounds:

The Huckleberry Finn case...illustrates a person who has demonstrably false beliefs about the good. But if virtue theorists insist that Huckleberry lacks virtue because of his flawed cognitive state, then this is bad news for most of us, who, even in some small way, are likely to harbor false views of value. The psychological requirements placed on virtue in the classical tradition seem far too rigid and unrealistic.

<sup>36</sup> Arpaly (2003), p. 78.

<sup>37</sup> See Holton (ms) for skepticism about the pervasiveness of such akratic action.

<sup>38</sup> I'm puzzled by Arpaly's take on this; she suggests that Huckleberry “finds himself” apologizing to Jim without having any moral insight or beliefs. See her (2003), p. 77. But a sincere apology expresses remorse; hence it requires the insight that one has treated someone in a way that one should not have. See Martin (2010).

Condition thus agrees that Huckleberry is a fundamentally good boy, even as it denies that he is morally praiseworthy for protecting Jim.

I will end by noting that this verdict fits rather well with Mark Twain's original story. Let's look at the crucial moment in which Huckleberry abandons his misguided plan to turn Jim in. Just as Huckleberry is departing from the raft that he and Jim share in order to tell on Jim, Jim calls after him:

Pooty soon I'll be a shout'n for joy, en I'll say, it's all on account o' Huck; I's a free man, en I couldn't ever ben free ef it hadn't been for Huck; Huck done it. Jim won't ever forgit you, Huck; you's de bes' fren' Jim's ever had; en you's de only fren' ole Jim's got now.<sup>39</sup>

When Huckleberry hears Jim's words, his resolve starts melting away:

I was paddling off, all in a sweat to tell on him; but when he says this, it seemed to kind of take the tuck all out of me.<sup>40</sup>

It seems that what changed Huckleberry's mind was Jim's reminder of their friendship. It's not surprising that this reminder should stop Huckleberry in his tracks. Huckleberry knows that loyalty is a central demand of friendship: friends stick together, don't tell on each other, and help each other in need.<sup>41</sup> And he cares about Jim's friendship. And so, it's not an accident that he acts like a good friend and protects Jim. But insofar as Huckleberry is motivated by considerations of friendship and he does not know that, in this situation, helping his friend is the morally right thing to do, it is a matter of luck that he does the right thing: after all, the demands of friendship may well occasionally conflict with moral demands.<sup>42</sup> As Twain tells the story, it seems plausible that we should regard Huckleberry as the beneficiary of moral luck: he is a fundamentally good boy who, when confronted with a moral situation that goes over his head, "accidentally does something good". The true hero here is Jim. By reminding Huckleberry of their friendship, Jim saves Huckleberry from committing a grave moral mistake.

### Bibliography

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

---

<sup>39</sup> Twain, p. 103.

<sup>40</sup> *ibid.*

<sup>41</sup> In fact, at a number of points in the novel Huckleberry explicitly reflects on these demands. My discussion here follows Manne (2013).

<sup>42</sup> See, for example, Cocking & Kennett (2000).



- Bennett, Jonathan (1974). The Conscience of Huckleberry Finn. *Philosophy* 49(188): 123–134.
- Cocking, Dean & Kennett, Jeanette (2000). Friendship and Moral Danger. *Journal of Philosophy* 97(5): 278–296.
- Dreier, James (2000). Dispositions and Fetishes: Externalist Models of Moral Motivation. *Philosophy and Phenomenological Research* 61(3): 619–638.
- Driver, Julia (2001). *Uneasy Virtue*. Cambridge University Press.
- Enoch, David (2014). In Defense of Moral Deference. *Journal of Philosophy* 111(5): 229–258.
- (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.
- Fricker, Miranda (2014). What's the Point of Blame? A Paradigm Based Explanation. *Noûs*. doi : 10.1111/nous.12067
- Gibbons, John. Knowledge In Action. *Philosophy and Phenomenological Research* 62(3): 579–600.
- Harman, Elizabeth (2011). Does Moral Ignorance Exculpate? *Ratio* 24 (4): 443–468.
- (forthcoming). The Moral Irrelevance of Moral Uncertainty.
- Hawley, Katherine (2003). Success and Knowledge-How. *American Philosophical Quarterly* 40 (1): 19–31.
- Hawthorne, John & Stanley, Jason (2008). Knowledge and Action. *Journal of Philosophy* 105 (10): 571–590.
- Henson, Richard (1979). What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action. *Philosophical Review* 88 (1): 39–54.
- Herman, Barbara (1993). *The Practice of Moral Judgment*. Harvard University Press.
- (1981). On the Value of Acting from the Motive of Duty. *Philosophical Review* 90 (3): 359–382.
- Hills, Alison (2009). Moral Testimony and Moral Epistemology. *Ethics* 120 (1): 94–127.
- Holton, Richard (2010). Norms and the Knobe Effect. *Analysis* 70 (3): 1–8.
- (ms). Inverse Akrasia and Weakness of Will.
- Hopkins, John (2007). What is Wrong with Moral Testimony? *Philosophy and Phenomenological Research* 74 (3): 611–634.
- Howell, Robert (2014). Google Morals, Virtue, and the Asymmetry of Deference. *Noûs* 48 (3): 389–415.
- Jones, Karen (1999). Second-Hand Moral Knowledge. *The Journal of Philosophy* 96 (2): 55–78.
- Kant, Immanuel (1785/1996). Groundwork of the Metaphysics of Morals, in: Gregor, Mary (trans. and ed.) *Practical Philosophy*. Cambridge University Press.

- Kearns, Stephen & Star, Daniel (2009). Reasons as Evidence. *Oxford Studies in Metaethics* 4: 215–42.
- Lillehammer, Hallvard (1997). Smith on Moral Fetishism. *Analysis* 57 (3):187–195.
- Manne, Kate (2013). On Being Social in Metaethics. *Oxford Studies in Metaethics* 8: 50–73.
- Markovits, Julia (2010). Acting for the Right Reasons. *Philosophical Review* 119 (2): 201–242.
- Martin, Adrienne M. (2010). Owing Up and Lowering Down: The Power of Apology. *Journal of Philosophy* 107 (10): 534–553.
- McGrath, Sarah (2011). Skepticism about Moral Expertise as a Puzzle for Moral Realism. *Journal of Philosophy* 108 (3).
- (2009). The Puzzle of Pure Moral Deference. *Philosophical Perspectives* 23 (1): 321–344.
- Nagel, Jennifer (2013). Knowledge as a Mental State. *Oxford Studies in Epistemology* 4: 275–310.
- Parfit, Derek (2011). *On What Matters*. Oxford University Press.
- Sher, George (2006). *In Praise of Blame*. Oxford University Press.
- Sliwa, Paulina (2012). In Defense of Moral Testimony. *Philosophical Studies* 158 (2): 175–195.
- Smith, Michael (1994). *The Moral Problem*. Blackwell.
- Stratton-Lake, Philip (2000). *Kant, Duty, and Moral Worth*. Routledge.
- Svavarsdottir, Sigrun (1999). Moral Cognitivism and Motivation. *Philosophical Review* 108 (2): 161–219.
- Twain, Mark (1884). *Huckleberry Finn*. Edited by Walter Blair and Victor Fisher (1988), University of California Press.
- Wallace, Jay (1996). *Responsibility and the Moral Sentiments*. Harvard University Press.
- Williams, Bernard (1981). Persons, Character, and Morality. In *Moral Luck*. Cambridge University Press.
- Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.