

WHAT'S THE MATTER WITH HUCK FINN?

HRISHIKESH JOSHI

ABSTRACT: This paper explores some key commitments of the idea that it can be rational to do what you believe you ought not to do. I suggest that there is a *prima facie* tension between this idea and certain plausible coherence constraints on rational agency.

I propose a way to resolve this tension. While akratic agents are always irrational, they are not always *practically* irrational, as many authors assume. Rather, “inverse” akratics like Huck Finn fail in a distinctively *theoretical* way. What explains why akratic agents are always either theoretically or practically irrational? I suggest that this is true because an agent’s total evidence determines *both* the beliefs *and* the intentions it’s rational for her to have. Moreover, an agent’s evidence does so in a way such that it’s never rational for the agent to at once believe that she ought to Φ and lack the intention to Φ .

1 Introduction

Akratic agents do what they believe they ought not to do. Through much of the history of ethics and the philosophy of action, it has been assumed that akratic agents are *irrational* in some sense: that there is *something* wrong with the state of mind of an agent who intends to do what, according to her own best judgment, she ought not to do. I agree.

Recently however, several authors, notably Nomy Arpaly and Robert Audi, have argued that it can be rational to act against your best judgment. It can, in some cases, be rational to do what you believe you ought not to do. For you can have a false belief as to what you ought to do, and in acting contrary to this belief, you can act for *good reasons*. In acting for good reasons, you act rationally. Arpaly (2003) presents us with several compelling cases in which it’s plausible that the akratic agents involved act rationally in this sense. Among

these akratics is Mark Twain's character, Huck Finn. Huck falsely believes he ought to turn in Jim, an escaped slave, to Miss Watson, the "owner." Yet, out of a concern for Jim's welfare, among other things, Huck chooses not to turn Jim in. Because Huck acts for good reasons, Arpaly contends that Huck acts rationally. And again, here, I agree.

But the ideas in the two paragraphs above seem to be in tension. On the one hand, there is the idea that there is something irrational about someone who acts akratically. On the other hand we have the claim that akratic Huck acts rationally. My aim is to resolve this apparent tension.

1.1 *The Importance of Acting Rationally*

Why care about whether an act is rational or irrational, as distinct from whether it is right, wrong, prudent, etc? The notion of acting rationally is philosophically interesting in its own right for the following reasons. First, our practice of giving and asking for *reasons* shows that we care not only *what* people do but also *why* they do it. Second, the notion of acting rationally is closely, and perhaps constitutively, related to the notion of acting *virtuously*. For many authors, to act virtuously is just to act in a practically rational manner. Thus, the notion of acting rationally is central to theories on which the idea of virtue is essential for understanding ethics.¹ Third, the concept of acting rationally encompasses the key notion of *moral worth*. While the concept of moral worth applies only when something of moral significance is at stake, the rational evaluation of actions can be appropriate more generally. In other words, a morally worthy action is just a rational action done in a context where the salient reasons are moral reasons.² Lastly, moral worth (and thus more broadly,

¹Plato and Aristotle were both very concerned with *virtue*, and Aristotle outlined a theory of what it is for a person to act virtuously. For Aristotle, practical wisdom (*phronēsis*) – in other words, practical rationality – is constitutive of virtue.

²The moral worth of an action, like the practical rationality of an action, depends on the reasons for which the action is done. If I save a drowning child merely so that I can be in the newspaper tomorrow, my action has low moral worth. By contrast, if I save the child out of a concern for her being able to live a full life, my action has higher moral worth. The concept of acting rationally is broader than the concept of moral worth because the former concept takes into account all the normative reasons an agent has, whereas the latter concept is only relevant when something of moral importance is at stake. Thus if I base my decision as to whether to work in the office or at home on a particular day upon the astrological profile of that day, my action is intuitively irrational – though (depending on what the astrological profile says) the action is neither morally worthy nor unworthy since nothing of moral importance is at stake.

rational action) is intimately connected to blameworthiness and praiseworthiness. Intuitively, actions with high moral worth are praiseworthy, and those with negative moral worth are blameworthy in the absence of excuses.

The concept of acting rationally, then, is a philosophically interesting one that bears important connections with other key concepts in ethics, such as moral worth and blameworthiness.

1.2 Roadmap

The plan for the paper is the following. In §2, I describe the case of Huck Finn in detail. I then formulate the central puzzle of the paper. In §3, I propose a principle which explains why a rational agent's ought beliefs will cohere with her intentions, but nonetheless is consistent with the claim that akratic agents can act rationally. I call this principle "evidence practicality." The idea is that an agent's total evidence determines *both* what it's rational for her to believe *and* what it's rational for her to do, in a way that precludes the possibility of it being rational for the agent to at once believe that she ought to Φ and fail to intend to Φ . In §4, I argue that evidence practicality explains the truth of the "wide-scope enkratic principle," defended by John Broome and others, according to which rationality requires an agent to intend to Φ if she believes she ought to Φ . The principle itself, however, will turn out not to be able to explain why any particular instances of akrasia exhibit failures of rationality, as many have thought. Rather, a correct explanation of why any particular instance of akrasia reveals a failure of rationality will consist of a *first-order, substantive* showing that the agent either improperly responded to her total evidence *or* she failed to be appropriately sensitive to the reasons she had, in forming her intention.

Immanuel Kant famously thought that the idea of moral worth was central to ethics. An action's moral worth, in Kant's words, lies "*not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon the realization of the object of the action but merely upon the *principle of volition* in accordance with which the action is done..." (G4:400) In contemporary philosophical parlance, an action's moral worth depends on the *reasons for which it's done*. The *good will*, out of which morally worthy actions spring, is for Kant the only thing in the world that "could be considered good without limitation". (G4:393) Indeed, according to Kant, "Even if, by a special disfavor of fortune or by the niggardly provision of a stepmotherly nature, this [good] will should wholly lack the capacity to carry out its purpose...then, like a jewel, it would still shine by itself, as something that has its full worth in itself." (G4:394) The concept of moral worth, then, lies at the heart of Kantian ethics.

In §5, I take up what is the obvious challenge to evidence practicality, namely whether the principle can satisfactorily account for the evidential role of moral testimony. One of the striking conclusions of this penultimate section will be that agents with sufficiently mistaken normative views *cannot* act rationally on the basis of normative testimony. Finally, in §6, I suggest that the resulting picture of rational action deflates the proper role of normative beliefs in guiding such action. However, I suggest that this is an interesting feature of the view, rather than a bug. I end with some exploratory remarks about what the proper role of normative beliefs could be within a rational agent's cognitive economy.

2 The Case of Huck Finn

Huck Finn is trying to decide whether to allow Jim to escape from Miss Watson. Huck believes, in keeping with the predominant moral outlook of his time, that Jim is Miss Watson's property. He also believes that it is one's moral duty to return lost property to the rightful owner. Huck thus has a false moral belief, namely, that he is morally required to turn Jim in to Miss Watson.³

Although Mark Twain's text does not explicitly say so, let's stipulate that Huck believes, not just that he's *morally obligated* to turn Jim in, but in addition, owing to the moral considerations involved, that this is what he ought to do *all things considered*. One might believe this if one thought that morality is *overriding* – that we always ought to do what morality requires us to do. But even if Huck doesn't think this, he might (perhaps implicitly) think that *in this situation*, the moral reasons he has to turn Jim in are sufficiently strong so as to make it the case that he all things considered ought to do so. Hence, let's suppose that, for one of these reasons Huck believes that he ought to turn Jim in, all things considered.

³The Huck Finn case has been employed in the moral psychology literature to defend a kind of “anti-intellectualism” about virtuous action. Bennett (1974) famously discusses this case to illuminate “misguided conscience.” Arpaly (2003) also talks about this example as an instance of what she calls “inverse” akrasia. My discussion of the case here doesn't aim to accurately characterize Huck's psychology as depicted in Mark Twain's novel – rather, I want to make certain stipulations about the scenario, and use it to illuminate more abstract claims about the purported requirement of enkrasia and the nature of evidence for normative beliefs.

However, Huck doesn't do this. Huck's act is thus formally akratic – he intentionally does what he believes he ought not to do.

Does Huck act *rationally*? The answer depends on features of the case we haven't specified. I'm going to suppose that the reasons for which Huck allows Jim to escape are good reasons. Since the moral reasons (as opposed to prudential or other kinds of reasons) are especially strong in this case, let's assume that Huck is moved chiefly by such reasons, whatever they may be. For example, if Kantianism is the right theory, let's suppose that Huck has come to appreciate that Jim is a rational being. Out of a respect for Jim's autonomy, Huck can't bring himself to turn Jim in. Or, if utilitarianism is true, let's suppose that Huck comes to see that happiness will be maximized if he allows Jim to escape. Huck realizes, say, that the gain in happiness for Jim if he secures his freedom is greater than the loss in happiness that Miss Watson will experience. As a result, Huck can't bring himself to turn Jim in.

So, does Huck act rationally? I want to consider three possible answers. The first is that Huck doesn't act rationally, because his act ought not to be done by his own lights – and it's irrational to do what you yourself believe you ought not to do.

The second possibility is that Huck's act is *partially* rational. On the one hand, Huck acts out of a sensitivity to the reasons that support his allowing Jim to escape. This fact counts in favor of the rationality of Huck's action. On the other hand, Huck's acting contrary to what he believes he ought to do, all-things-considered, detracts from the rationality of his action.

The third possibility is that Huck's act is *fully* rational. His action reflects his sensitivity to the considerations that in fact bear on what to do in his situation – and this suffices to make his act rational. His false ought judgment is irrelevant; it does not affect the rationality of his action.⁴

⁴According to some authors, for instance Broome (1999), rationality consists at least partly in satisfying *normative requirements*, where doing so can come apart from acting for the right reasons. These authors would presumably want to endorse one of the first two options mentioned here if they think that the enkratic requirement is a requirement of rationality. Nonetheless, as I go on to show, one can capture the intuitive claim that the enkratic requirement is a requirement of rationality even if one thinks that rationality is a matter of being properly responsive to reasons. However, this will mean giving up the idea that the enkratic requirement is explanatorily *basic* in a certain sense. This paper thus suggests a way in which the enkratic requirement can fit in within a picture – defended in, for example, Dancy (2002), Arpaly (2003), Schroeder

Now, the plausibility of each of these answers will depend on what we mean by someone's acting rationally. The concept of acting rationally I'm interested in is one that plays a central role in ethics and decision theory. An agent acts rationally in this sense just in case the *practical reasoning* leading up to it is *good*, where practical reasoning is reasoning that concludes in an action or intention.

What is good practical reasoning? Is it just reasoning that leads the agent to do what she ought to do, or one of the things she may do (if there is no unique required act)? No. Someone can do what she ought to do, or what she may do, as a result of good practical reasoning, or as the result of terrible practical reasoning. Consider the following possibilities: 1) Huck frees Jim because he thinks that doing so will allow Huck to blackmail Jim in the future; 2) Huck frees Jim because it will require too much effort to report Jim to Miss Watson; and 3) Huck frees Jim simply because it's a Tuesday. In all these cases, Huck's practical reasoning is patently bad even though it leads Huck to do what he ought to do. When I ask whether Huck's act is rational, I'm not simply asking whether the act is what he ought to do – rather, I'm asking whether there is anything wrong with the practical reasoning that leads him to choose the act.

I want to examine the commitments involved in endorsing the third option mentioned above. That is:

RATIONAL HUCK: Huck acts rationally in allowing Jim to escape.

Arpaly (2003) defends the view that formally akratic acts can be rational by appealing to the case of Huck Finn, along with a plethora of other compelling examples.⁵ My aim in this paper is not to further defend the claim. Rather, I want to examine whether it can be made consistent with certain very plausible coherence constraints on rational agency.

(2009), Markovits (2010), and Parfit (2011) – on which acting rationally (or acting well, or acting virtuously) consists in properly responding to reasons.

⁵Others who argue for the view include Audi (1990) and Weatherston (2013).

2.1 *The Central Puzzle and a Sketch of the Solution*

The problem is that there seems to be a tension between RATIONAL HUCK and the following eminently plausible idea:

RATIONAL COHERENCE: A state of mind containing an intention to Φ in circumstance C, as well as a belief that one ought *not* to Φ in C, is to some extent irrational.

RATIONAL COHERENCE is the idea that *something* is going wrong with the agent, like Huck, who forms the intention to do what he believes he ought not to do. Indeed, it seems that if an agent is ideally rational, then his intentions will match his beliefs as to what he ought to do.

Prima facie, RATIONAL COHERENCE and RATIONAL HUCK do not sit well together. The first thesis says that something is going wrong with Huck if his act is formally akratic. On the other hand, RATIONAL HUCK says that Huck's formally akratic act is rational.

I want to find a way out of this prima facie tension. My hope will be to find an explanation of why RATIONAL COHERENCE is true which does not conflict with RATIONAL HUCK. My preferred explanation will draw a sharp distinction between theoretical and practical rationality. I will argue that Huck is theoretically *irrational* though he is practically rational. On the other hand, mundane cases of akrasia, of the sort widely discussed in the literature on weakness of will, exhibit *practical irrationality* on part of the agents involved, but not theoretical irrationality.

Thus it will turn out that formally akratic agents are indeed failing in some respect *qua* rational. Hence, RATIONAL COHERENCE is true. However, I shall contend that the explanation for why such agents exhibit irrationality is not, as many have thought, that they are violating the formal constraint given by:

ENKRATIC PRINCIPLE: Rationality requires (if you believe you ought to Φ , you intend to Φ).⁶

⁶This principle is defended as a “wide-scope” requirement of rationality in Broome (1999), among others.

Rather, akratic agents fail to be fully rational because they fail to properly respond to certain first-order reasons that they have.

Run-of-the-mill akratics fail to be appropriately sensitive to certain practical reasons for action, and are irrational for that reason (and not because they violate ENKRATIC PRINCIPLE). Inverse akratic agents like Huck fail to be appropriately sensitive to certain theoretical reasons for forming ought beliefs, and are irrational for *that* reason. On this picture, though ENKRATIC PRINCIPLE may well be *true*, it is not explanatory. No one is irrational *because* she acts akratically. Rather, some akratic agents are irrational in a distinctively practical way, others in a distinctively theoretical way. ENKRATIC PRINCIPLE is thus at best a corollary of the correct view of these matters, and not, as it is commonly believed, an explanatory principle in its own right.

3 Resolving the Tension

Notice that RATIONAL HUCK only says that Huck's *act* is rational; it is silent on the rationality of his ought belief, namely that he ought to turn Jim in. If this belief is irrational, then RATIONAL COHERENCE is satisfied – for then, we are able to say that *something* is wrong with Huck's state of mind insofar as his decision to free Jim is concerned.

Hence, my proposal for understanding what's going on in the Huck Finn case described above is the following. Huck's act, or alternatively, his intention to free Jim, is rational. Hence RATIONAL HUCK is satisfied. However, his belief that he ought not allow Jim to escape is *irrational*. Generalizing this result leads to the following principle:

DISJUNCTIVE IRRATIONALITY: Whenever an agent believes that she ought to Φ but does not have the intention to Φ , her belief that she ought to Φ is *theoretically irrational* or her failure to have the intention to Φ is *practically irrational*.

Below I suggest an even more basic principle that undergirds DISJUNCTIVE IRRATIONALITY.

3.1 *The Practical Import of Evidence*

What is it for a belief to be irrational? Plausibly, it is for the belief to be such that one's total evidence does not support having it. For instance, it's irrational for me to believe that the earth is flat. Why? Because I have a plethora of evidence that counts against the claim that the earth is flat – for example, I have seen pictures of the earth from space, and the earth's being roughly spherical explains why there are seasons, tides, etc.⁷ Moreover, the evidence that could potentially support the belief – testimony from the ancient past, perception of the ground as flat, etc. – is defeated. In light of my total evidence, then, it would be irrational for me to believe that the earth is flat.⁸ Importantly, the rationality of my belief depends on how it is based on the evidence I actually have – not what I *take* to be evidence for the belief in question.⁹

If a belief is to be rational, it needs to be a proper response to the agent's total evidence. Now suppose the belief on part of an agent A, that she ought to Φ is rational. According to DISJUNCTIVE IRRATIONALITY, then, her not intending to Φ would constitute practical irrationality. So it must be the case that her having the evidence she has, evidence that makes it rational for her to believe that she ought to Φ , also somehow makes it the case that she is practically irrational in not intending to Φ . The following principle suggests itself:

EVIDENCE PRACTICALITY: For any agent A, with total evidence E, E makes
it theoretically rational for A to hold the belief that A ought to Φ if and only

⁷Generally, it is thought that if P is part of the best explanation for why Q is true, then Q is evidence for P.

⁸According to Kelly (2014),

It is characteristic of rational thinkers to respect their evidence. Insofar as one is rational, one is disposed to respond appropriately to one's evidence: at any given time, one's views accurately reflect the character of one's evidence at that time, and one's views manifest a characteristic sensitivity or responsiveness to changes in one's evidence through time. Of course, rationality is no guarantee of correctness. Indeed, in a given case one might be led astray by following one's evidence, as when one's evidence is *misleading*. But being mistaken is not the same as being unreasonable. (§2)

⁹In this paper, I want to remain neutral with respect to different conceptions of evidence in epistemology – however the distinction between an agent's evidence and what the agent takes to be her evidence is available for all plausible views of evidence. For example, a currently popular view – see Williamson (2000) – holds that agent's total evidence just consists of all the things she *knows*. If this view is right, then an agent's evidence will amount to what she knows, rather than what she thinks she knows.

if having E makes it the case that A would be practically irrational in not intending to Φ .

This, I contend, is a deeper explanation for why DISJUNCTIVE IRRATIONALITY is true. The idea is that the theoretical rationality of a normative belief, like that of other beliefs, depends on the agent's evidence – but so does the *practical rationality* of the agent's intentions.

3.2 Evidence Practicality vs. the Standard Model

EVIDENCE PRACTICALITY is in effect a rejection of what I call the Standard Model, as depicted in Figure 1. On this model, practical rationality is a matter of having one's intentions or actions correspond to one's ought beliefs. The practically rational agent is one whose ought beliefs are effective in guiding her actions.¹⁰

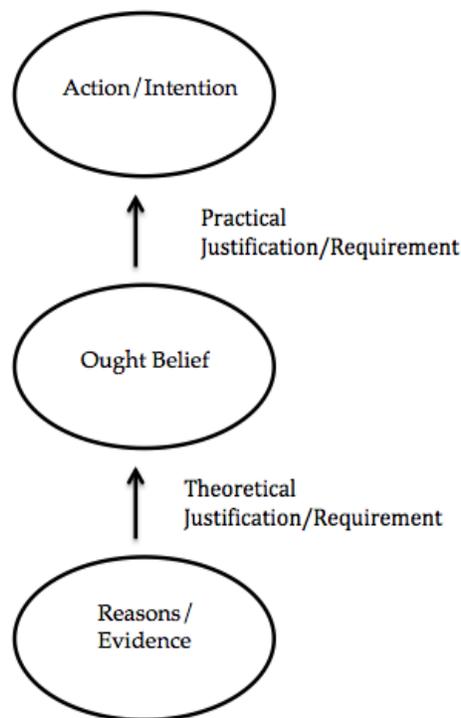


FIGURE 1. Standard Model

Later on, I will discuss what I think is the main motivation behind the implicit acceptance of the Standard Model by several authors. For now, I want to point out briefly that the

¹⁰The Standard Model is implicit in the arguments of many authors writing on rational action and akrasia. For a recent, explicit defense of this idea, see Smith (2013).

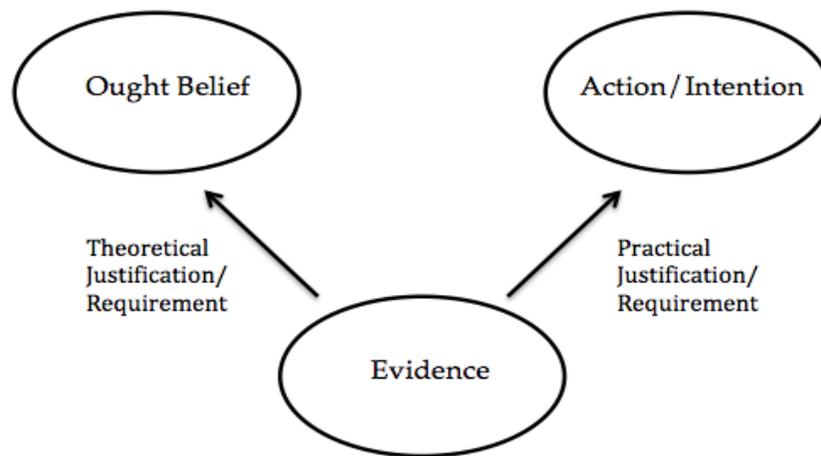


FIGURE 2. Evidence Practicality: A Sketch

Standard Model is problematic for two main, independent reasons. First, it seems that young children can exhibit practical rationality, in the sense that they can properly respond to their reasons – but it’s an open question whether they have ought beliefs in any robust sense. Second, it seems implausible that *every* practically rational action is one that is caused in the right way by the agent’s ought beliefs. An agent might do something in a rational manner, but nonetheless might not have given much conscious thought to it – in such cases, it seems to be a stretch to characterize the agent as having had the appropriate ought belief.

EVIDENCE PRACTICALITY is a departure from this model. As Figure 2 illustrates, evidence, on this picture, can play a *direct justificatory* role with regards to actions or intentions, without having to go through the channel of an agent’s ought beliefs. Figure 2 is intended as a broad sketch of EVIDENCE PRACTICALITY; below, I suggest a way in which *reasons* can be brought into the picture.

3.3 *Reasons and Evidence*

The theoretical rationality of a belief depends on the believer’s total evidence and on the way in which the belief is sensitive to (i.e. based on) the relevant evidence. The concept of *evidence* then plays this key role in epistemology, with regards to our appraisals of the rationality of beliefs.

What plays the analogous role with regards to our appraisals of the rationality of *actions*? It is common in the present literature to assume that the rationality of an action has to do with the *reasons* for which it is done. The rationality of an intention similarly depends on the reasons for which it is held. What are reasons? Reasons for action are simply considerations that count in favor of performing an action.

What it's rational for an agent to do depends on reasons, I contend, but not on the reasons *there are* but rather the reasons the agent *has*. This is an important distinction when it comes to appraising an agent's actions vis-à-vis their rationality. The basic idea is that in order to *have* a reason, the agent must possess some epistemic access to that reason. There can be all sorts of reasons to which we have no epistemic access – but intuitively, these won't affect the rationality of our acting one way or another.

To make the contrast stark, let's suppose there is, unbeknownst to me, a blue button in the room in which I'm working, which, if pressed, will drastically reduce world poverty. Now if this is true, there is a (very strong) reason for me to press the button – doing so will improve so many lives! However, my failure to press this button will reveal no moral or rational failure on my part – and that's because I don't have the proper epistemic access to this fact. I don't know about this button, nor am I in a position to know, nor am I justified in believing that there is such a button. In this sense, I don't *have* a reason to press the button.¹¹

The idea that the practical rationality of an action is determined by the reasons the agent has, along with EVIDENCE PRACTICALITY, then yields that the *reasons an agent has are determined by her total evidence*. Though I want to say that the reasons an agent has are determined by her total evidence, I don't want to *identify* reasons with evidence. Some authors do identify these two things. According to Kearns and Star (2009), a reason to Φ is just evidence that one ought to Φ . However, this thesis is stronger than I need for this paper. Furthermore, this claim invites a style of objection wherein a case is presented in which it's plausible that some fact F is evidence that A ought to Φ but isn't a reason for A

¹¹For more discussion of this distinction and its philosophical implications, see Lord (2015).

to Φ . Thus, all I want to claim in this paper is that given that an agent A is in circumstance C, A's total evidence will fix the reasons she has.

As Figure 3 illustrates, the resulting picture is this. Given an agent A in circumstance C, A's total evidence E will determine what beliefs it's rational for A to have, and a fortiori, what *normative* beliefs it's rational for A to have. This is the theoretical side of the picture. On the practical side, E will determine the reasons A has, which in turn will determine what intentions it's rational for A to have/what actions it's rational for A to choose. The way in which evidence ultimately determines both the theoretical rationality of beliefs and the practical rationality of intentions has the following striking feature. It's *never* the case that, given agent A and circumstances C, A's total evidence E at once makes it theoretically rational for A to believe that she ought not to Φ *and* provides her with reasons which make it the case that she would be practically rational in Φ -ing.

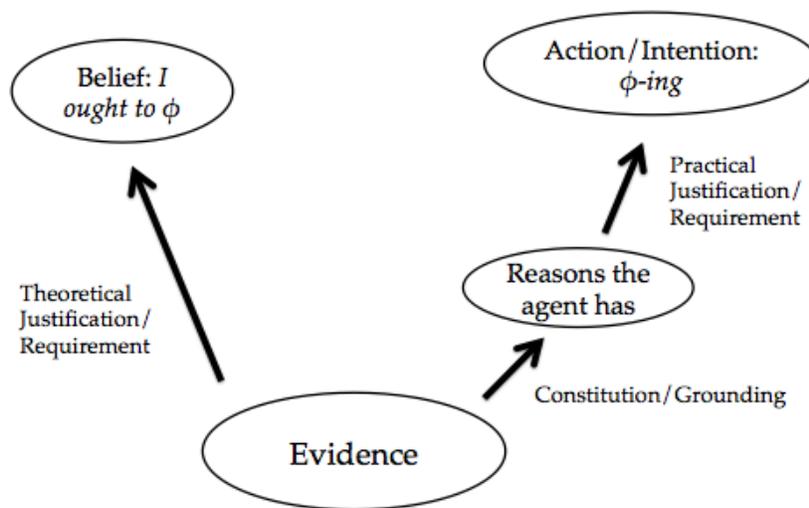


FIGURE 3. Evidence, beliefs, and actions

4 A New Picture of Akrasia and the Enkratic Requirement

DISJUNCTIVE IRRATIONALITY says that all instances of akrasia involve some irrationality. It thus offers a way of vindicating a conviction that an overwhelming majority of philosophers

have had, namely that akrasia constitutes a failure of rationality, at least to some extent. Any given akratic act either reveals practical irrationality or theoretical irrationality or both. The view, however, abandons the assumption that all instances of akrasia are failures of *practical* rationality.

I'm proposing DISJUNCTIVE IRRATIONALITY and EVIDENCE PRACTICALITY as mid-level principles. They are mid-level in the sense that they are explanatorily more basic than RATIONAL COHERENCE – they explain this latter principle; yet, given a particular instance of akrasia, the explanation of why that instance reveals an irrational state of mind won't bottom out in either of the mid-level principles. Rather, such an explanation will have to show either that the relevant ought-belief is irrationally held, or that the intention is not properly sensitive to the reasons the agent has, or both. These explanations will essentially be first-order, substantive explanations.

So what's the matter with Huck Finn? The problem is that his ought-belief – viz., that he ought to turn Jim in – is not one that Huck's evidence makes it rational for him to have. Notice that Huck has plenty of evidence that supports the claim that Huck ought to help Jim to escape – by coming to know Jim, Huck has come to see that Jim is a person like himself, with his own desires, aims, likings and dislikings. This evidence plausibly supports the claim that Jim's welfare matters, which in turn supports the claim that (other things equal) he ought to allow Jim to escape. I don't mean to suggest that this is all the relevant evidence – but this is the sort of evidence that I think Huck's ought-belief is insufficiently sensitive to. This then, is the beginning of what I call a *first-order* or *substantive* explanation of why Huck's formally akratic state of mind is irrational in some sense.

4.1 *Run-of-the-mill Akrasia*

In contrast with the Huck Finn case, run-of-the-mill cases of akrasia, I think, do exhibit practical irrationality. In these cases, agents show insufficient responsiveness to the reasons they have.

Here's an example. Bob knows he ought not to smoke. He has read the relevant bits of

the research showing the causal relationship between smoking and cancer and a host of other ailments. He knows that smoking decreases life expectancy. In light of this, he believes he ought not to smoke. Yet, it's Friday night, and he has had a few drinks. He sees a group of friends smoking outside the bar, goes out, and has a cigarette – all the while having the ought belief regarding smoking.

Bob's ought belief is rational – it's a proper response to the evidence he has pertaining to the causal relationship between smoking and various illnesses. Yet, when Bob has had a couple of drinks, the reasons he has to avoid smoking lose their grip on his deliberation. Bob smokes despite having strong reasons not to – reasons to do with avoiding cancer and other diseases. His act is practically irrational in this way; indeed, it is a platitude that alcohol can weaken reason's control over one's actions.

4.2 *Whither the Enkratic Requirement?*

As discussed above, my favored explanations for why formally akratic acts exhibit irrationality differ depending on the case. In cases of run-of-the-mill akrasia, where the agent has a true and justified belief that she ought to Φ but does not Φ , what's going on is that she's practically irrational – the reasons she has do not guide her deliberations in the proper way. On the contrary, in cases of “inverse” akrasia, where the agent has a false belief that she ought to Φ , but still ends up Φ -ing for the right reasons, the agent's ought-belief is not just false, but also irrationally held. Inverse akratics like Huck exhibit theoretical rather than practical rationality.

If this picture is right, there is a real asymmetry between the two kinds of akrasia – they exhibit different kinds of irrationality. In other words, they are irrational for different kinds of reasons. Therefore, there is no one requirement that such agents violate that explains why their states of mind are irrational.

Consider the “wide-scope” enkratic requirement defended by John Broome, among others:

ENKRATIC PRINCIPLE: Rationality requires (if you believe you ought to Φ ,

you intend to Φ).¹²

This, I contend, has no role to play in the explanation for why akratics are irrational to some extent. First, notice that neither of the two explanations rehearsed above appealed to this principle. Second, because ENKRATIC PRINCIPLE is “wide-scope” it can be satisfied in two ways in any given case: if you believe you ought to Φ and don’t intend to Φ , you can satisfy the principle by either revising your belief or your intention. Now, Huck is in a situation where he believes he ought to turn Jim in but does not intend to do so. But notice that Huck can satisfy the enkratic requirement by revising his intention, and deciding to turn Jim in. But this, according to my picture, would make him no more rational.¹³ In fact, it would make him *more irrational*. For now, not only would he be responding to his evidence in a deficient way, but he would also be improperly responsive to the reasons he has to allow Jim to escape!

Many authors seem to assume that ENKRATIC PRINCIPLE is not only true, but *explanatorily basic* in the sense that certain states of mind are irrational fundamentally *because* they constitute a violation of the enkratic requirement.¹⁴ In other words, the assumption is that when there is a mismatch between an agent’s belief as to what she ought to do and her intentions, the mismatch is the *source* of her irrationality. According to the picture presented here, such explanatory claims are mistaken – the source of her irrationality is never *this* mismatch. Rather, the source of her irrationality has to do either with the mismatch between her evidence and her ought beliefs or the mismatch between her evidence (and thus

¹²See Broome (1999) and Broome (2013). In the latter, a more detailed formulation is given. Broome writes:

Rationality requires you to intend what you believe you ought; it requires you not to be akratic.

More accurately:

Enkrasia: Rationality requires of N that, if

- (1) N believes at t that she herself ought that p , and if
- (2) N believes at t that, if she herself were then to intend that p , because of that, p would be so, and if
- (3) N believes at t that, if she herself were not then to intend that p , because of that, p would not be so, then
- (4) N intends at t that p . (p170)

¹³See Arpaly (2003) for a more thorough defense of this claim.

¹⁴For example, see Broome (2013), Davidson (1980), and Wedgwood (2007).

her reasons for action) and her intentions.

However, I do not want to deny that ENKRATIC PRINCIPLE is *true*. It's true because of the way that an agent's evidence determines what she should believe, as well as what intentions it's rational for her to have. In other words, ENKRATIC PRINCIPLE is true, and it's true because EVIDENCE PRACTICALITY is true.

5 Evidence Practicality and Testimony

EVIDENCE PRACTICALITY is the core of the solution offered in this paper. In my opinion, it can explain two things: 1) why Huck's *act* of freeing Jim is rational and 2) why there is *something* irrational about an agent's state of mind when he does what he believes he ought not to do. However, it is a somewhat striking claim: according to EVIDENCE PRACTICALITY, it's *never* the case that an agent's total evidence E at once supports her believing that she ought to Φ while also making it practically rational for her to not Φ .

A preliminary worry with this idea is that perhaps EVIDENCE PRACTICALITY is not really plausible in the case of Huck. For, perhaps since Huck has plenty of *testimony* – owing to the place and time in which he lives – to the effect that Jim is Miss Watson's rightful property, it's not irrational for Huck to form the belief that he ought to turn Jim in. This is what people have told Huck all his life and he is rational in listening to them, even if this leads him to hold mistaken beliefs. Nonetheless, we might think, Huck's act of allowing Jim to escape is rational because it is responsive to the reasons he has – reasons to promote Jim's welfare, and the like. If this way of construing the Huck Finn case is right, then EVIDENCE PRACTICALITY is false.

My response to this worry is two-fold. First, I think that Huck's testimony-based evidence to the effect that Jim is Miss Watson's rightful property is *defeated*. It's defeated partly by the information that Huck has gathered through his interactions with Jim; given that Jim's psychology is sufficiently similar to Huck's, in the morally relevant ways, the case to be made for differential assignment of rights is radically undermined.

Second, our granting that Huck's belief is not the ideally rational response to the evidence he possesses need not commit us to having to *blame* Huck for his belief. Indeed, Huck plausibly has several excuses. For one, he's just a child. What is more, he lives in a place and time where the predominant moral outlook has it that some people can be owned by others. Given the costs of holding beliefs that contradict the predominant moral outlook of one's time it may be unfair, in some cases, for us to blame agents who align their moral beliefs with those of their culture, even if those beliefs are not justified. In any case, I suspect that some of the pull towards claiming that Huck's false moral belief is epistemically justified owes its force to the desire not to blame Huck for his false belief. But these two things can come apart.

This two-fold response is the one I favor vis à vis the Huck Finn case. However, it's not the only way to understand the case in a manner compatible with EVIDENCE PRACTICALITY. If it turns out that Huck's evidence does support his belief that he ought to turn Jim in, then all we have to say in order to maintain EVIDENCE PRACTICALITY is that Huck's act is *irrational*, even if it's what he ought to do. For it would then turn out that in freeing Jim, Huck is not properly responsive to the reasons he has. Indeed, it seems to me that the more it's made plausible that Huck's evidence militates in favor of his ought-belief, the less plausible it is that Huck acts rationally.

A second preliminary worry is whether EVIDENCE PRACTICALITY makes it either impossible or too hard for agents to have *rationaly held* but *false* normative beliefs. Huck, I've been arguing, has false normative beliefs, but they are irrationally held. Indeed, according to EVIDENCE PRACTICALITY, *whenever* an agent can be practically rational in Φ -ing on the basis of the reasons he has, it will be irrational for him to believe that he ought not to Φ . Nonetheless, surely it seems possible and rather common for agents to have false but justified normative beliefs. If EVIDENCE PRACTICALITY cannot accommodate this datum, it seems to involve implausibly strong commitments.

However, EVIDENCE PRACTICALITY is consistent with the thesis that we can rationally hold false normative beliefs. In fact, it's fairly easy to be justifiably mistaken about lots

of normative claims. But the reasons for this are quite mundane. Here is an illustration. Suppose I have every reason to believe that the white powder in the jar is regular sugar. The jar is labeled ‘sugar,’ it’s in a household cabinet, and the white stuff has a granular texture. You ask for some sugar in your tea. Consequently, I form the following belief: I ought to put this stuff in your tea. This belief is rational for me to form even if, unbeknownst to me, a nefarious neighbor has laced the stuff with arsenic. The mundane reason, then, why we can have plenty of rationally held but false normative beliefs is simply that the non-normative evidence on which they rest can be misleading.

5.1 *Acting on the Basis of Normative Testimony*

The following sort of case is trickier. The case involves testimony regarding a normative matter from a normative expert. Suppose it can be rational to act on the basis of such testimony. The following question arises for a defender of EVIDENCE PRACTICALITY: what is the evidence gained through such testimony which makes it rational for the agent to do what the expert says the agent ought to do, by providing the agent with good reasons to do the act in question? Below, I describe how I think a defender of EVIDENCE PRACTICALITY should answer this question. The answer I defend will turn out to have a striking implication: it will turn out that people whose normative views are sufficiently mistaken *cannot* rationally follow normative testimony.

Consider the following case. Suppose Bob is faced with the choice of pushing a particular red button. All Bob knows is that the button is hooked up to some complex and powerful device – so that whether or not he presses the button may well have momentous consequences. Now suppose Oracle is a normative oracle. That is, she knows all the normative facts there are to know. Furthermore, Bob knows that Oracle is a normative oracle. Now, Oracle tells Bob, “you ought to press that button.” We have stipulated that Bob is rational in believing that Oracle is a normative oracle (he *knows* this); it then follows that Bob is rational in believing that he ought to press the red button. Can he be practically rational in pressing the button? I said earlier that practical rationality is a matter of properly responding to the

reasons one has. What, exactly, are the reasons that Bob has to press the red button, after Oracle has made her pronouncement?

As an aside, it is an open question whether we can ever be justified in believing in “moral experts” or “normative experts” who are reliable on moral/normative matters in the way that, say, accountants are (generally) reliable on taxation matters, and physicists are reliable sources of knowledge about physics. Perhaps there is a deep disanalogy here between the normative and non-normative domains. But for the purposes of this paper, I assume that it is possible to know that someone is a moral/normative expert. (If this turns out to be false, all the better for my purposes – for the worry being considered does not arise.)¹⁵

Waiving such skepticism about the possibility of knowing whether someone is a normative expert, the puzzle we are faced with is this. Bob knows that he ought to press the button. This raises the following question.

Question: What are the reasons for which Bob can act, so that his act is practically rational?

As emphasized earlier, being practically rational is *not* simply a matter of doing what you ought to do. You have to do what you ought to do by responding to the right reasons. Below are three possible answers to the question posed above. I think we should definitely reject the first one. The second answer is more tenable in my opinion, but the third answer is the best.

Answer 1: Bob comes to know that he ought to press the red button. He can act in a practically rational way by simply acting for the reason that *he ought to press the red button*.

Problem: There are several issues with this line of response; presented here are two main worries. First, on many conceptions of reasons, reasons are explanations.¹⁶ In particular, reasons are explanations for ought-claims. But if that’s right, and if the fact that A ought to Φ can itself be a reason to Φ , then it would seem that the fact that A ought to Φ explains

¹⁵See Hills (2009) and McGrath (2011) for more discussion of this and related issues.

¹⁶Authors of diverse metaethical commitments have defended this view. See, for example, Dancy (2002) and Schroeder (2007).

itself! This is problematic.¹⁷ An analogous issue arises for views of reasons on which reasons are evidence for ought-claims.¹⁸ The problem there is that we would have to say something can be evidence for itself.

Second, the response gives rise to the “moral fetishism” worry, *à la* Williams (1982) and Smith (1994). The worry is that intuitively, the virtuous agent should care about the morally relevant features of his action, rather than caring about doing something simply because it’s right *de dicto*. So if Bob presses the button simply for the reason that he ought to do so, he’d be acting out of a fetish for morality (or in this case, normativity in general).

The issues raised for Answer 1 can be resolved if we adopt:

Answer 2: Bob’s reason for acting is: *Oracle said that I ought to press the red button*. This claim is a *non-normative* claim because the ought clause occurs in a disquotational context. The worries about the possibility of a fact explaining itself, or a claim being evidence for itself do not arise.

Problem: The chief problem with this proposal is that it’s hard to see how the fact that Oracle said Bob ought to press the button can itself be a *reason* for Bob to press the button. Consider again the view of reasons as explanations. It’s implausible that the Oracle’s merely *saying* that Bob ought to press the button *explains* why Bob ought to press the button. Rather, suppose that the button is hooked up to a complex system, which, when operational will help mitigate climate change; the button starts the system and Bob is the only one in a position to press the button. If this is true, then it seems that what explains why Bob ought to press the button is the following fact: Bob’s pressing the button will help mitigate climate change. Therefore, if Bob is to act for good reasons, then, he can’t act simply for the reason that Oracle said that he ought to press the red button.

Answer 3: Depending on Bob’s background views on morality and the extent to which they are justified, Bob will come to possess non-normative evidence about the properties of the button, and can act rationally on the basis of reasons given to him by this evidence. Thus suppose for simplicity that act-utilitarianism is true and that Bob believes that it’s

¹⁷Markovits (2010) also points out this problem.

¹⁸Notably, Thomson (2008) defends a view of reasons as evidence for ought-claims.

true. Hence, the balance of aggregate pleasure over pain, or net happiness, is the only morally relevant feature of the world. Moreover, the ways in which one's actions affect net happiness entirely determines what one ought to do. Supposing these things, when Oracle tells Bob that he ought to press the red button, Bob gains the following evidence: pressing the red button will maximize happiness. Thus, Bob can act for the reason that his pressing the button will maximize happiness. And in this way, Bob can act *rationally*, i.e. for the right reasons – he is not doomed to act fetishistically as described above.

According to Answer 3, then, an agent with *true normative beliefs* can act rationally in following pure normative testimony (more precisely, the degree to which an agent will be able to rationally follow normative testimony will depend on the extent to which his fundamental moral beliefs are true). For, such an agent will be in a position to infer that acting in accordance with the normative testimony will constitute doing what he ought to do *de re*, and not just *de dicto* – Bob, in the example above is able to infer that pressing the button will maximize happiness and can act on *this* basis, rather than being forced to press the button simply because that's what he ought to do.

Act-utilitarianism is a simplifying assumption above, but it's easy to extend the picture to cover more complex moral theories. Suppose, for instance, that the normative domain consists solely of constraints – α , β , and γ . The normative commandments, as it were, are: do not violate α ; do not violate β ; and do not violate γ . Applying this stipulation to the case above, Bob will be able to infer from Oracle's testimony that his failing to press the button will constitute his violating constraint α , constraint β , or constraint γ . Thus, Bob will be able to act rationally in choosing to press the button – he can act for the reason that failing to press the button would constitute violating one or more of these constraints.

Bob in the example above has the *correct* normative views, and thus is able to act rationally in following normative testimony. What if someone had *false* normative views? Thus suppose Bill believes that the sole thing he ought to do is to spread the message of Cult, which happens to be a morally bankrupt organization. When Oracle tells Bill that he ought to press the button, what would Bill infer from this? Bill would infer that his pressing the button would

spread the message of Cult, and can act for this reason. But that's a terrible reason! His act would be a vicious, irrational one. What would make his act less vicious and irrational is if he presses the button simply because he ought to do so or simply because Oracle said that he ought to do so. While this makes Bill's act less irrational, it's nonetheless still irrational owing to the arguments rehearsed above, in Answers 1 and 2.

Thus, a person like Bill is doomed to act irrationally when acting on the basis of normative testimony. Huck is similarly doomed (but perhaps less so, since he surely has *some* true normative beliefs). Of course, this is not to say these agents cannot act rationally, period. Indeed, they can choose to act on reasons which by their own lights are bad reasons. Huck does this when he frees Jim. But neither Huck nor Bill can act rationally simply on the basis of normative testimony.

6 The (Purported) Role of Normative Beliefs in Guiding Action

If I'm right in denying that ENKRATIC PRINCIPLE *explains* why akratic states of mind exhibit irrationality, then I'm in effect deflating the role of *normative beliefs* in guiding action. However this is, I think, a (positive) feature rather than a bug of the resulting picture.

Part of the appeal of the opposing view – i.e. the Standard Model of Figure 1 – is perhaps that it allows us to make sense of what separates rationally guided action from mere bodily movements as well as *arational* actions. If you apply pressure to my arm so that it goes up, my arm's going up is merely a bodily movement, as opposed to an action of mine. And when I absent-mindedly twiddle my thumbs, I may well be acting in some sense, but my action is not rationally guided. We might also think that acts done from compulsive urges, whims, and addictions are not rationally guided.

But what separates such arational actions and bodily movements from rationally guided actions? One answer might be: for something to be a rationally guided action, the agent's *normative beliefs* have to play a certain role in guiding the behavior. In other words, the agent's normative beliefs have to control her actions in the right kind of way. If this answer

is right, then when an agent's actions do not comport with her normative beliefs, we have an instance of practical irrationality – an agent acting akratically is like an agent acting from some compulsive urge.

This is one answer, but not the only possible answer. I think we should distinguish between rationally guided action and action where the agent's normative beliefs play a guiding role. The two are not equivalent. Rationally guided action, in my view, is action done for *reasons*. And one can act for reasons even if one acts contrary to what one believes one ought to do. Indeed, I think akratic agents can act for reasons; Huck, for instance, acts for reasons – good ones, in fact.

With this distinction at hand, there is a simple, flat-footed answer to the question as to what separates rationally guided actions on the one hand and arational actions or bodily movements on the other. The answer is: rationally guided actions are those *done for reasons*. When you apply pressure to my arm and it goes up, I don't raise my arm for a reason. Similarly, I (usually) don't twiddle my thumbs for a reason.

Hence we should distinguish *reason* being in the driver's seat from *normative beliefs* being in the driver's seat. Once this distinction is made, the case for the *explanatory role* of the ENKRATIC PRINCIPLE is undermined.

That said, the picture I defend does raise some interesting questions. What is the proper role of normative beliefs in the rational agent's cognitive landscape, if not in guiding her behavior? What is the point of doing ethics, if all that matters in acting virtuously is acting for the right reasons? I want to take up these questions elsewhere.

But I do think that there are (obviously) important roles that normative beliefs play in our cognitive lives. In part, they allow us to *scrutinize* our reasoning and behavior. Second, they seem to play an essential function in enabling us to *justify* our conduct to others. However, these two roles of normative beliefs have been underemphasized, in my opinion, and as a result the purported role of normative beliefs in guiding behavior has seemed to be paramount.

7 Conclusion

The view described here is able to resolve the apparent tension between two claims: 1) it's possible to act rationally against your best judgment; and 2) there is *something* irrational about at once believing you ought to Φ while at the same time not intending to Φ . Thus, a defender of 1) may accept the account presented here, rather than being forced to reject 2), which is an eminently plausible claim.

According to the account presented here, akrasia always involves irrationality but does not always involve *practical irrationality*, as many have thought. An agent can act rationally against her best judgment – i.e. her belief as to what she ought to do. However, such an agent's belief as to what she ought to do *has to be irrational*. This is because the practical rationality of her actions, as well as the theoretical rationality of her beliefs, depends on her *evidence*. And if the evidence gives her good reasons to Φ , it cannot at the same time justify her believing that she ought not to Φ .

Regular akratics and “inverse” akratics are irrational for different reasons. The former suffer from practical irrationality, while the latter suffer from theoretical irrationality. Thus the enkratic principle, which requires that an agent do what she believe she ought to do, is not explanatory – it doesn't explain why any particular case of formal akrasia involves irrationality. Rather, the enkratic principle is a true consequence of something more basic – namely the way in which an agent's evidence determines both what it's rational for an agent to do and what it's rational for her to believe.

References

- Aristotle. *Nicomachean Ethics (Translated, with Introduction, Notes, and Glossary, by Terence Irwin)*. Hackett Publishing Company, Inc., second edition, 1999.
- Nomy Arpaly. *Unprincipled Virtue*. Oxford University Press, 2003.

- Robert Audi. Weakness of will and rational action. *Australasian Journal of Philosophy*, 68 (3):270–281, 1990.
- Jonathan Bennett. The conscience of huckleberry finn. *Philosophy*, 49:123–34, 1974.
- John Broome. Normative requirements. *Ratio*, 12:398–419, 1999.
- John Broome. *Rationality Through Reasoning*. Wiley-Blackwell, 2013.
- Jonathan Dancy. *Practical Reality*. Oxford University Press, 2002.
- Donald Davidson. How is weakness of the will possible? In *Essays on Actions and Events*. Oxford University Press, 1980.
- Alison Hills. Moral testimony and moral epistemology. *Ethics*, 120(1):94–127, 2009.
- Immanuel Kant. *Groundwork of the Metaphysics of Morals*. Cambridge University Press, 1997.
- Stephen Kearns and Daniel Star. Reasons as evidence. *Oxford Studies in Metaethics*, 4: 215–42, 2009.
- Thomas Kelly. Evidence. *Stanford Encyclopedia of Philosophy*, 2014.
- Errol Lord. Acting for the right reasons, abilities, and obligation. *Oxford Studies in Metaethics*, 10, 2015.
- Julia Markovits. Acting for the right reasons. *Philosophical Review*, 119(2):201–242, 2010.
- Sarah McGrath. Skepticism about moral expertise as a puzzle for moral realism. *Journal of Philosophy*, 108(3), 2011.
- Derek Parfit. *On What Matters*. Oxford University Press, 2011.
- Plato. *Complete Works*. Hackett Publishing Company, Inc., 1997.
- Mark Schroeder. *Slaves of the Passions*. Oxford University Press, 2007.
- Mark Schroeder. Buck-passers' negative thesis. *Philosophical Explorations*, 12(3), 2009.
- Michael Smith. *The Moral Problem*. Wiley-Blackwell, 1994.

Michael Smith. The ideal of orthonomous action, or the how and why of buck-passing.

In Brad Hooker David Bakhurst and Margaret Olivia Little, editors, *Thinking about Reasons: Themes from the Philosophy of Jonathan Dancy*. Oxford University Press, 2013.

Judith Jarvis Thomson. *Normativity*. Open Court, 2008.

Brian Weatherson. Running risks morally. *Philosophical Studies*, 167(1):1–23, 2013.

Ralph Wedgwood. *The Nature of Normativity*. Oxford University Press, 2007.

Bernard Williams. *Moral Luck*. Cambridge University Press, 1982.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.