From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Philosophy

2016

Avoiding the One Thought Too Many: Essays on Virtuous Agency

Joshi, Hrishikesh Suhas

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01f1881p38j

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

AVOIDING THE ONE THOUGHT TOO MANY: ESSAYS ON VIRTUOUS AGENCY

Hrishikesh Suhas Joshi

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
PHILOSOPHY

Advisers: Michael Smith, Gideon Rosen, and Philip Pettit

September 2016

© Copyright by Hrishikesh Suhas Joshi, 2016. All rights reserved.

Dissertation Abstract

This dissertation is about the role of moral beliefs within virtuous agency. I argue that this role has been overestimated – having, and being guided by, the correct moral beliefs is neither necessary nor sufficient for virtuous action. Chapter 2 argues that certain agents can act rationally even if they act contrary to their beliefs about what they ought to do. The main task of this chapter is then to accommodate the platitude that there is *something* irrational going on in an akratic agent.

Virtuous agents, I suggest, are those in whom the normative structure of the reasons they have is mirrored in the motivational structure they exhibit. Key to this picture is the notion of *taking there to be a reason*, which I analyze in terms of dispositions to *reason*. The task of Chapter 3 is to give an account of reasoning (or inference), which eschews the ideology of taking there to be a reason, thus challenging the dominant strand in the literature.

The fourth chapter gives an account of the subjective ought – the ought that is tied to practical rationality and praise/blame. According to some authors, an agent ought to ϕ in this sense just in case she believes she ought to ϕ. Others argue that only non-moral beliefs are relevant to determining what one ought to do in this sense. I argue for the claim that the agent's beliefs, be they moral or non-moral, are *irrelevant* – what matters is the agent's *total evidence*. Furthermore, I argue, what an agent subjectively ought to do is what her maximally virtuous counterpart (but importantly, one who is *not idealized* in other ways) with the same total evidence would do given her circumstances.

Acknowledgments

First and foremost, I would like to thank my three advisors, Philip Pettit, Gideon Rosen, and Michael Smith, for their unhesitant support and guidance throughout the dissertation process and graduate school in general. I could not have asked for advisors who were more generous with their time and resources. I am also indebted to Elizabeth Harman, Thomas Kelly, Sarah-Jane Leslie, Sarah McGrath, and David Plunkett for reading and providing very helpful feedback on earlier chapter drafts. Marcus Gibson, Sherif Girgis, and Samuel Preston gave valuable comments through the Princeton Philosophy Department's writing sample review program.

Talking through some of the core ideas presented here has been enormously helpful, and in that regard, special thanks go to: Daniel Berntson, Arlyss Gease, Keinan Greene, Mark Harris, Christopher Howard, Eric Hubble, and Noel Swanson. Furthermore, this dissertation has benefited tremendously from comments given by audiences at: the University of Michigan at Ann Arbor, the Pacific APA, the Rocky Mountain Ethics Congress, and Princeton University. In this regard, I want to thank Johann Frick and Boris Kment for particularly helpful suggestions. Shamik Dasgupta's feedback has been crucial in helping frame some of the key questions that this dissertation aims to answer. I would also like to thank David Copp, Hanti Lin, and Domenica Romagni for useful discussion on some of the topics involved in this project.

I want to reserve particular thanks for Gary Watson, and especially Scott Soames, for guiding and mentoring me through my transition into professional philosophy. Without their training and encouragement, I would not be in a position to write a philosophy dissertation at an institution like Princeton.

And lastly, I want to express gratitude to my parents and family. Most especially, thank you Meghan and Henry for your constant moral support, and for putting up with me as I worked to complete this project.

Contents

Dissertation Abstract iii

Acknowledgments iv

1 Introduction **1**

2 What's the Matter With Huck Finn? **11**

2.1 Introduction 11

2.1.1 The Importance of Acting Rationally 12

2.1.2 Roadmap 13

2.2 The Case of Huck Finn 14

2.2.1 The Central Puzzle and a Sketch of the Solution 17

2.3 Resolving the Tension 18

2.3.1 The Practical Import of Evidence 19

2.3.2 Evidence Practicality vs. the Standard Model 20

2.3.3 Reasons and Evidence 21

2.4 A New Picture of Akrasia and the Enkratic Requirement 23

2.4.1 Run-of-the-mill Akrasia 25

2.4.2 Whither the Enkratic Requirement? 25

2.5 Evidence Practicality and Testimony 27

2.5.1 Acting on the Basis of Normative Testimony 30

2.6 The (Purported) Role of Normative Beliefs in Guiding Action 34
2.7 Conclusion 36

3 Dispositionalism about Reasoning **37**
3.1 Introduction 37
3.2 The Problem of Inference 38
3.2.1 The Antecedent Disposition Thesis 40
3.3 Dispositionalism 42
3.3.1 The Intuitive Sketch 42
3.3.2 The Details 44
3.3.3 Dispositions and Masking 50
3.3.4 Impossible Inferences and Logical Entailment 52
3.4 Broome's Rule-Following Proposal 53
3.4.1 The Problem 54
3.4.2 Possible Responses 55
3.5 Causation in Virtue of Rationalization 57
3.5.1 The Problem 58
3.6 The Doxastic Proposal 60
3.6.1 The Over-Intellectualization Problem 61
3.6.2 A Dilemma for the Doxastic Proposal 62
3.7 Rejecting the Taking Condition 64
3.8 Practical Reasoning 65
3.8.1 Robustness of Intention 66
3.8.2 Other Kinds of Reasoning? 67
3.9 Conclusion 68

4 Evidence and Obligation **70**

4.1 Introduction 70
4.2 Two Notions of Obligation 70
4.3 Why Moral Beliefs Do Not Matter 72
4.4 Why Non-Moral Beliefs Do Not Matter Either 74
4.5 Evidentialism About Obligation 80
4.6 Reframing the Debate Between Uncertaintists and Externalists 85
4.7 From Evidence to Action: Ideal vs Non-Ideal 88
4.7.1 Towards a Virtue-Theoretic Conception of the Subjective Ought . . . 90
4.7.2 A Problem and Two Solutions for Virtue-Based Analyses of Obligation 93
4.7.3 Reasons-Based Accounts of the Subjective Ought 95
4.8 Conclusion 97

Chapter 1

Introduction

This introductory chapter aims to make a substantive case in motivating the work to follow in later chapters, and describes the way in which the resulting picture is meant to hang together.

What role should an agent's moral beliefs play in guiding her practical reasoning? Only a derivative one, I suggest. In this dissertation, I argue that an agent's doing what she *believes* to be right, just, obligatory, etc. does not add to the quality of, or the virtue displayed by, her action. Rather, the virtuous agent is one who is properly responsive to her reasons.

The bulk of the dissertation is devoted to motivating as well as working out the details and commitments of such a picture. Along the way, I explore how the picture affects some key recent debates concerning: i) what an agent should do under conditions of moral uncertainty, ii) the propriety of deferring to moral testimony, and iii) whether moral ignorance can be exculpatory.

I begin by considering the case of Huck Finn. In Mark Twain's book, Huck is traveling with Jim, an escaped slave. Huck believes, in keeping with the moral outlook of his time, that he ought to turn Jim in to Miss Watson, the "owner." However, Huck can't bring himself to do this. As I set up the case, we are to imagine that Huck acts out of a concern for Jim's welfare and a respect for his humanity.

What are we to say about Huck's action? Is he acting as he ought to? In one sense, the answer is obviously: yes. Slavery is morally abhorrent, and so Huck is doing the right thing in letting Jim go. This is the objective perspective on the situation. This is what an ideally rational (perhaps we have to add: moral) agent would want Huck to do.

But there is another sense of 'ought' that we might be interested in. This is the sense in which it would be a *mistake* not to do as you ought to. Non-mistaken action and the objectively correct action can come apart in obvious ways. A doctor in the 6th century BC who takes great pains to deliver a particular treatment to his patient may be not be making a mistake, though as a matter of fact, the treatment is ineffective. He couldn't have known any better! Another way to put it is: he acts as he *subjectively* ought to, but not as he *objectively* ought to (supposing some other available remedy works better).

This sense of ought is also plausibly connected to *blame* – along with other authors, I think it forms a necessary condition on the propriety of blame. That is, an agent can only be blameworthy for Φ-ing if she ought not have Φ-ed in the subjective sense.[1]

Now, is Huck doing as he ought in this *latter* sense? I contend: yes. And this is so even if he believes he ought not to free Jim. I would also add, following Arpaly (2003) that he acts in a practically rational manner – for he is properly sensitive to the morally relevant considerations at hand. He acts out of a respect for Jim's humanity and a concern for his welfare. Huck displays virtue in freeing Jim.

If this claim is right, it raises a puzzle which I address in the first chapter. The puzzle is this. It seems truistic that there is *something* irrational about an agent who at once believes he ought to Φ, but does not intend to Φ. In other words, there is something irrational about an akratic state of mind. But if Arpaly is right, Huck acts in a practically rational manner. So where can we locate his irrationality?

I propose in Chapter 2 that we should separate theoretical and practical rationality. Doing so allows us to say that while Huck's act is practically rational, it is nonetheless the-

[1]See, for example, Harman (2015) and Milo (1984) for further discussion and defense of this idea.

oretically *irrational*. Thus, Huck's akratic state of mind does involve irrationality. However, I claim, it does not involve *practical* irrationality. The temptation to think that Huck's act is practically irrational comes from the widely held assumption that akrasia must involve practical irrationality.[2]

Instead, I claim that in any given instance of akrasia either the agent's belief that she ought to Φ is theoretically irrational *or* her intention not to Φ is irrational (or both). What explains why this is true? Why should every akratic act exhibit irrationality in one or both of these ways?

I suggest that we can make progress on this question by noting that theoretical rationality is tied to an agent's total evidence. A theoretically rational belief is one that is a proper response to an agent's total evidence. An agent's total evidence thus constrains what it's rational for her to believe. So, if practical rationality is to covary with theoretical rationality so that akrasia always ends up being irrational, then practical rationality must be tethered to total evidence as well.

I thus put forth a principle I call Evidence Practicality, according to which 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 such a way that it's never fully rational to at once believe that one ought to Φ, but intend to not Φ. That is, for any agent A and total evidence E, E makes it theoretically rational for A to believe that she ought to Φ if and only if E also makes it practically rational for A to Φ. The rest of the chapter takes up some important challenges to Evidence Practicality.

Now, in order for Huck's act to be practically rational, he has to be responsive to the appropriate reasons bearing upon the action of allowing Jim to escape. Indeed, in order to be practically rational in doing some action, it's not enough simply to perform the action which is in fact supported by the reasons. Huck's allowing Jim to escape is in fact the action that is supported by the balance of reasons. But if Huck allows Jim to escape merely so

[2]See, for example, Smith (1994).

that he can blackmail Jim in the future, or merely because it's a Tuesday, then, intuitively, Huck's action would not be practically rational or virtuous.

What does it take for an action to be practically rational? Markovits (2010) suggests the agent's *motivating* reasons behind the action must coincide with the *normative* reasons there are for the agent to perform the action. In other words, the agent's reasons for doing the action must match the reasons that in fact *justify* the action.

While Markovits puts her thesis in terms of moral worth – she offers the above condition as necessary and sufficient for an act to have full moral worth – the lesson presumably generalizes. Whereas the concept of 'moral worth' is only properly applied to actions where something morally weighty hangs in the balance, 'practical rationality' as I use the term, takes into account reasons in general. Thus, it can be practically rational for one to avoid procrastinating and work on a paper for the reason that it will add to one's publication record – though in usual circumstances, this act plausibly does not merit the label *morally worthy*. This is because the reasons at play are presumably non-moral (unless one has made a promise to finish the paper, or the like).

There is a problem with Markovits' analysis, however. And this is that reasons can function in different ways, and plausibly the practically rational actor will register these differences in her motivational structure. As Gert (2012) has pointed out, normative reasons can have at least two dimensions: *justifying* and *requiring*. For instance, the fact that giving Mary some flowers on her birthday will make her happy is a justifying reason for giving her flowers. However, the fact that you promised to get her flowers on her birthday is a requiring reason to do so. However, Gert observes, motivating reasons can have *only one* dimension of strength. Therefore, Gert argues that Markovits' analysis fails because of the mismatch in dimensionality, as it were, between normative and motivating reasons.

In light of this problem, I think that we need the notion of an agent's *taking* something to be a reason in order to give an adequate account of practically rational action. An agent can take something to be a reason in more than one way. To directly respond to Gert's

worry, an agent can take R to be a reason to Φ in a requiring way *or* a justifying way.

To take something to be a reason for an action, I suggest, is to be disposed to *deliberate* in certain ways. Thus, for instance, I take the fact that giving Mary flowers will make her happy as a *justifying* reason to give her flowers if the following is true. In deliberation, I am generally disposed to weigh this consideration against others – for example, the fact that flowers cost money, take time to procure, etc. And in the absence of counterweights (or defeaters), I am disposed to reason from the fact that giving Mary flowers will make her happy to the intention to give her flowers. In this sense, the fact that giving Mary flowers will make her happy factors as a positive consideration within the deliberative context of deciding whether to give her flowers.

Taking something to be a reason in a *requiring* way will look different. Suppose I have promised Mary to get her flowers on her birthday. I take this to be a requiring reason if the following is true: I am disposed to *block* certain considerations from functioning as counterweights (in the way described above) in the context deciding whether to get Mary flowers. Thus suppose that the flowers cost $10 and will take 15 minutes to acquire. I will not be disposed to use these considerations as counterweights in my reasoning. To put it differently, the fact that I promised Mary to get her flowers *defeats* these other considerations from functioning as countervailing reasons, within the context of my reasoning.

A good analogy is perhaps a judge's order to have certain evidence be made inadmissible in court. The judge's doing so blocks that evidence from figuring in the jury's deliberation. The jury then will not use that evidence in deciding the verdict. Likewise, in the example above, considerations of cost and time (up to some threshold) are blocked off from figuring in the deliberation as to whether to get flowers for Mary. This is just what it is to treat the promise as generating a requiring reason.

Given the notion of taking there to be a reason, we are able to give a more satisfactory account of practical rationality. For an act to be rational, *the normative structure of the reasons the agent has for doing a particular action must match the deliberative structure*

of the agent. That is, she must take the requiring reasons to be requiring reasons, the justifying reasons to be justifying reasons, and so on. In addition, the deliberation that leads to a particular decision or intentional action must in fact be correct.

What I am suggesting here, then, is that i) we need the notion of 'taking something to be a reason' in order to give a satisfactory account of practical rationality (and hence moral worth) and ii) we should analyze this notion in terms of *dispositions to reason* in certain ways. *Taking something to be a reason*, then, is to be analyzed in terms of *reasoning*.

Chapter 3 takes up the task of giving an account of reasoning. The central problem is to separate genuine reasoning from mere associative thinking. The problem is in some sense analogous to the problem of action as posed by Wittgenstein, when he wrote, "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"[3] The analogous question is: what is left over if we subtract the fact that in A's head, the thought P was followed by the thought Q from the fact that A *reasoned* from P to Q? Alternatively, what must we add to the fact that the thought P was followed by the thought Q to get the fact that the agent reasoned from P to Q?

It's neither necessary nor sufficient that P entails or supports Q . For, we can make invalid or bad inferences. And there is nothing to stop associative thinking from leading to Q from P where P entails or supports Q. Neither is it enough to say the thought P causes the thought Q – this is true of associative thinking as well. The main contender in the literature, defended in Boghossian (2014) and others, is that A reasons from P to Q if her taking P to support Q plays the proper causal role in getting her to transition to Q. This is, however, the opposite of what I am suggesting – I suggest that we analyze taking P to support (i.e. be a reason in favor of believing/doing) Q in terms of dispositions to reason.

In the chapter, I give independent reasons to think that Boghossian's account is wrong. One such reason is the following. Consider Archimedes when he first discovers the Principle of Buoyancy. Let P = *the water in the bathtub rose to extent X* and let Q = *the Principle*

[3]Wittgenstein (1953), §621

of Buoyancy is true. I suggest that it's implausible to describe Archimedes as *antecedently* having possessed the mental state of *taking P to support Q*. That's precisely why the episode counted as a *discovery*!

The account I offer avoids this problem and captures the fact that Archimedes' inference is a discovery. On the view defended in Chapter 3, A reasons from P to Q just in case the thought that P causes the thought that Q and simultaneously causes A to acquire the disposition to reduce confidence in Q upon reducing confidence in P. Additionally, the causation involved must be direct. The basic idea is that through reasoning, we come to base our conclusions upon our premises.

Following the description of this account, the rest of the chapter is devoted to discussing objections and to explaining how the account avoids some key problems that I argue other prominent accounts face – in particular, those presented in Broome (2013) and Arpaly and Schroeder (2013).

Another key positive feature of analyzing 'taking there to be a reason' in terms of dispositions to reason, and in turn analyzing reasoning in the way sketched above is that the picture separates taking there to be a reason from *believing* there to be a reason. I take this to be crucial for several reasons.

First, consider the case of Huck Finn again. Here, Huck is taking Jim's welfare and his humanity as reasons for allowing him to go free – that's why we consider his act to be morally worthy. But it's plausible that Huck doesn't *believe* that these are reasons to do what he does. Moreover, even if Huck does believe that these are reasons to let Jim escape, he plausibly only believes they are *justifying* reasons, which are defeated by the property rights considerations of Miss Watson. If this is right, taking can't be believing – the reasons Huck takes there to be, as well as the extent to and manner in which he takes these considerations to be reasons, come apart from Huck's *beliefs* about reasons.

The point generalizes. To borrow another example from Arpaly, the committed believer in Ayn Rand's philosophy may believe that she does not have reasons to help others, unless it

serves her own interests in some way, but nonetheless may display great concern for the well being of others. And such genuine concern involves not merely certain emotional dispositions but also entails that the agent takes others' welfare to generate reasons for action.

Second, thinking of taking as believing in this context leads to a hitherto unnoticed problem with regards to reasons created by *valuing*. It seems plausible that your valuing something can create reasons for you with respect to that thing, over and above the extent to which it is *valuable*. Thus your valuing the opera more than the theater can make it the case that you have more reasons to donate to the opera as opposed to the theater – even if we grant that both activities are equally valuable. This extra reason you get for donating to the opera, I claim, owes its existence to the fact that you *value* the opera more.

Of course, someone might resist this initial intuition by saying that it's not your valuing *per se* that generates the extra reason – it's merely that since giving to the opera will give you greater satisfaction, you ought to give to the opera. This worry can be rebuffed by isolating the factor(s) that it cites. So, think of a case where a nefarious neurosurgeon has set things up so that once you click the donation button and decide which cause to donate to, she will cause you to forget that you donated and forget that you had the money to give in the first place. It still seems that you have more reason to give to the opera – you care more about it!

If all this is granted, a problem arises for the taking = believing view when we consider what the attitude of valuing consists in. Plausibly, valuing something consists in more than just emotional dispositions towards that thing. According to the most prominent accounts of valuing – see for example, Scheffler (2010) and Frankfurt (2004) – valuing consists in part in taking there to be reasons to act in certain ways. The other component of valuing involves possessing certain emotional dispositions, the nature of which will differ depending on the particular object of valuing.

To illustrate: valuing an art-form might consist in part in taking there to be reasons to go see it if it's displayed nearby, taking there to be reasons to preserve it, and so on. Valuing a

relationship might consist in taking there to be reasons to spend time with the other person, to promote their welfare, and so on. Indeed, it's hard to imagine a case where A values O, but O figures in no way in her deliberative landscape.

But now there is a problem. If valuing consists (in part) in taking there to be reasons for acting in certain ways, and if taking = believing, then it would seem that believing R to be a reason to Φ can, in part, *make it the case* that R in fact is a reason to Φ. This seems problematic insofar as we think of belief as a *representational* state of mind. Believing something cannot make it so![4]

There is a third reason to prefer my analysis over a view on which taking = believing, and reasoning is to be understood in terms of taking there to be reasons. As I explain in more detail in Chapter 3, it seems that young children can reason, but it's not obvious that they possess the concept of a reason. But believing that R is a reason to Φ presumably requires possession of the concept.[5]

Now, as mentioned earlier, in Chapter 2, I suggest Evidence Practicality, if true, can guarantee that all instances of akrasia contain some irrationality. One of the two key components of that principle is that what an agent ought to do in the subjective sense (and thus, what it's practically rational for her to do) is determined by the agent's *total evidence*. Chapter 4 provides independent support for this contention.

The main alternatives in the literature are (to a first approximation) that the subjective ought is determined by i) what the agent believes; ii) by the agent's *non-moral* beliefs only, and iii) only the agent's *justified* beliefs. i) and ii) form the basis of a live debate between "uncertaintists" and "externalists" respectively. I give arguments for why all three of these views are unsatisfactory.

[4]There seem to be some cases where the belief that P can partly ground the fact that P. For example, my believing that there are beliefs partly grounds the fact that there are beliefs, if Rosen (2010) is right that existential facts are grounded in their instances. I submit however, that the case at hand above is importantly different.

[5]Boghossian (2014) is also worried about this issue, and thus ends up rejecting the taking = believing view. Ultimately, he doesn't offer an account of taking, and posits it as a primitive.

What an agent subjectively ought to do, I contend, is determined by a function from the agent's circumstances and her *total* evidence. Nonetheless, I do think that the important kernel of the disagreement between uncertaintists and externalists can be captured within the total evidence view, and I explain this in detail in the middle sections of Chapter 4.

Finally, I address the question of what the function from circumstances and total evidence to action looks like. Take an ordinary agent A in circumstances C and total evidence E. What ought A subjectively to do? Is it what the ideally rational agent would do, or would want A to do, given C and E? I say: no. Doing so marks a major departure from a dominant strand in metaethical theory with regards to the construction of "ideal observers."

Too much idealization leads to crucial problems, which I examine towards the end of Chapter 4. Ultimately, I defend a view on which, given an important proviso, what A ought to do is given by what her ideally *virtuous* but *not* ideally *cognitively sophisticated* counterpart, with evidence E, would do in C. In this sense, then, virtue might turn out to be prior to obligation.

Chapter 2

What's the Matter With Huck Finn?

2.1 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.

2.1.1 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,

[1]Plato 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.

[2]The 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.

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

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.

2.1.2 Roadmap

The plan for the chapter is the following. In §2.2, I describe the case of Huck Finn in detail. I then formulate the central puzzle of the paper. In §2.3, 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 §2.4, 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.

In §2.5, 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

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.

normative views *cannot* act rationally on the basis of normative testimony. Finally, in §2.6, 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.

2.2 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.

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

[3]The 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.

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]

[4]According 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.

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

[5]Others who argue for the view include Audi (1990) and Weatherson (2013).

2.2.1 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]

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.

2.3 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*

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

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.

2.3.1 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

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

[8]According 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)

[9]In 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.

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 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.

2.3.2 Evidence Practicality vs. the Standard Model

EVIDENCE PRACTICALITY is in effect a rejection of what I call the Standard Model, as depicted in Figure 2.1. 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]

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 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.

[10]The 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).

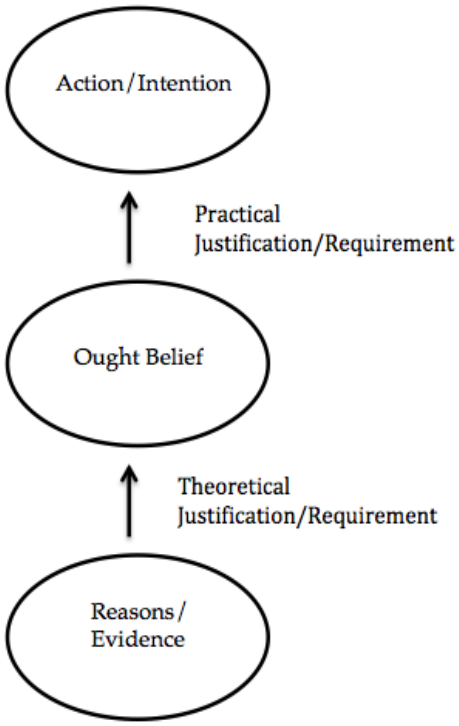


Figure 2.1: Standard Model

EVIDENCE PRACTICALITY is a departure from this model. As Figure 2.2 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 2.2 is intended as a broad sketch of EVIDENCE PRACTICALITY; below, I suggest a way in which *reasons* can be brought into the picture.

2.3.3 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

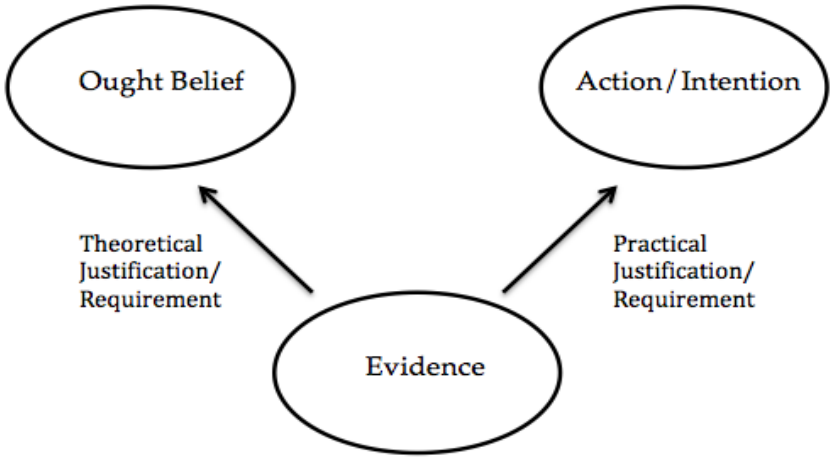


Figure 2.2: Evidence Practicality: A Sketch

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 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 2.3 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.

2.4 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

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

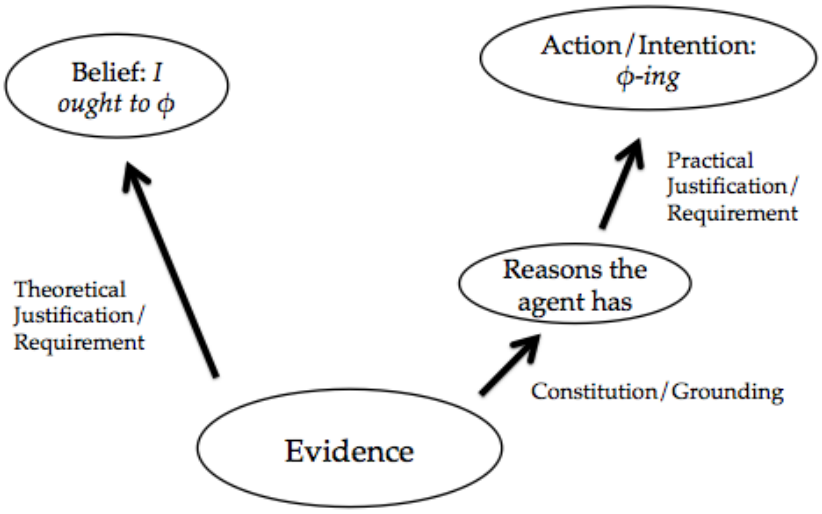


Figure 2.3: Evidence, beliefs, and actions

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.

2.4.1 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.

2.4.2 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 irrationality.

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

[12]See 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)

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 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.

2.5 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

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

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

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 *rationally 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.

2.5.1 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 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

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

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

[17]Markovits (2010) also points out this problem.

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

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 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.

2.6 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.

2.7 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.

Chapter 3

Dispositionalism about Reasoning

3.1 Introduction

This chapter gives an account of reasoning that distinguishes it from mere associative thinking. I defend two main claims. First, I argue that the best account of reasoning will be *dispositional*. I do so in part by rejecting the three main proposals available in the recent literature: 1) John Broome's rule-following account of reasoning, 2) The Causation in Virtue of Rationalization Proposal, defended in Arpaly and Schroeder (2013) and Wedgwood (2006), and 3) the Doxastic Proposal on which normative beliefs – to the effect that the premises support the conclusion – are essentially involved in reasoning.

Second, I argue that the best version of a dispositional account of reasoning involves rejecting what Boghossian (2014) has called the "taking condition." The taking condition says that reasoning, say from P to Q, essentially involves one's *taking P to support Q* and thereby concluding Q *because* of this taking. I contend that the taking condition cannot be accommodated within a plausible dispositional account of reasoning. Rather, the dispositions corresponding to the reasoner's "taking P to support Q" are *acquired through* the process of reasoning, rather than being a *precondition* of it.

3.2 The Problem of Inference

Imagine a thinker, Alice, going through the following series of thoughts, in direct succession:[1]

ASSOCIATIVE THINKING: 1) Saul Kripke is a philosopher. 2) London is pretty.

INFERENCE: 1) The S&P 500 went up by more than 2% today. 2) The Fed kept rates steady.

Let's suppose that in the latter case, but not the former, Alice performs an *inference*; she engages in a bit of reasoning, she draws a conclusion. She *infers* that the Fed kept rates steady from the fact that the S&P 500 went up by more than 2%. To put it differently, she *reasons from* 1) to 2). But what *makes it the case* that the latter is an inference but the former is not?

Logical entailment won't help us here. 2) is not entailed by 1) in either case. If we want to insist on logical entailment as a necessary condition for reasoning, we will have to insist that one cannot, strictly speaking, perform the bit of reasoning described in INFERENCE. Rather, INFERENCE only describes Alice's reasoning enthymematically. That is, when Alice infers 2) from 1), she does so only via: 1.5) If the S&P 500 went up by more than 2% today, then the Fed kept rates steady.

But since reasoning is a mental process that we can perform consciously, the natural test for whether 1.5) is a necessary thought for Alice to perform the reasoning detailed in INFERENCE is simply to see whether we have any such thoughts whenever we perform everyday inferences. And the answer is clearly: no. When I sometimes see the car in the driveway and reason that somebody is home, I have never caught myself thinking, "If there is a car in the driveway, then someone is home," and I doubt that this is idiosyncratic. Perhaps, one may think, my implicitly holding such a belief is what renders my thinking reasoning. I will consider and reject this proposal later on.

[1]In this paper, I use 'thought' to mean occurrent belief or supposition; a thing thought need not be believed – it may be merely supposed.

For now, there is an even simpler reason why the logical entailment condition is too strong. And this is that we clearly can perform *bad* inferences, which flout the rules of logic. We can affirm the consequent, for example.

For the same reason, a weaker notion of support won't help either. In INFERENCE, one might think, 1) inductively supports 2). But one can perform the inference detailed above even if in fact 1) does not support 2). Maybe in recent history, the market has been able to predict the Fed's actions and thus would not fluctuate in response to them. Someone who does not realize this, however, can still perform INFERENCE. Thus, the requirement that entailment or some kind of evidential support must hold between the claims involved within a process of reasoning is too strong.

In another way, the condition is *too weak*. For one can think the thoughts in INFERENCE *without* performing an inference, even supposing that 1) supports 2). In reading the case detailed in INFERENCE above, you probably had the thoughts with the contents given by 1) and 2) but you most likely didn't perform an inference *from* 1) to 2). Clearly, some kind of causal connection is required between your thinking 1) and your thinking 2). Moreover, this causal connection needs to be direct, or proximate. If Alice's occurrently believing 1) leads her to check the financial news, where she reads that (and thus comes to believe that) 2) is the case, she hasn't thereby reasoned from 1) to 2).[2]

Nonetheless, a causal connection between thoughts, however direct, will not suffice to make it the case that a mental process amounts to the performance of an inference. Consider the following way of filling out ASSOCIATIVE REASONING. Alice has recently read "A Puzzle about Belief," written by Saul Kripke, within which the sentence 'London is pretty' appears several times. Owing to this, when the belief that Saul Kripke is a philosopher becomes occurrent in her mind, Alice is reminded of London's being pretty. So, her occurrent belief

[2]Occurrent beliefs are at the forefront of one's mind, in some sense. You believe that the Earth revolves around the Sun, but before reading this sentence just now, your belief was not occurrent. So, just before you read the previous sentence, we may say you had the "dispositional" but not occurrent belief that the Earth revolves around the Sun. I will say a bit more about this later on. For further explication of this distinction, see Schwitzgebel (2010).

that Saul Kripke is a philosopher *causes* her to have the occurrent belief that London is pretty. However she does not infer that London is pretty from the claim that Saul Kripke is a philosopher (though this bizarre process of reasoning could in principle occur).

So here is where we are. A thinker's reasoning from P to Q must involve her thought that P directly causing her thought that Q. What more do we need to add to give an account of reasoning?

3.2.1 The Antecedent Disposition Thesis

An initially plausible idea is that when one reasons from P to Q one's coming to think Q must be explained by some *antecedent disposition*. That is, one must be antecedently disposed to come to think Q upon thinking P, and this disposition must play the appropriate causal role in one's coming to think Q. Perhaps that is what is missing in ASSOCIATIVE THINKING but present in INFERENCE?

Unfortunately, there are several problems with this Antecedent Disposition Thesis.

Firstly, there is the problem of specifying the *stimulus conditions* for the dispositions in question. When one reasons from P to Q, the natural suggestion would be that the stimulus condition is the thought that P.

However, one can, in general, infer from some P not only to some Q, but to a myriad of other claims – R, S, and so on. In INFERENCE, Alice can infer from 1) to claims like: Bob sold some of his shares; the FTSE went up too; and so on. In fact, one's thinking P may only cause one's thinking Q once in a lifetime – perhaps one out of a thousand times one has thought P. It's odd to say then, that one has the disposition to think Q upon thinking P.

Another possibility is that the stimulus condition involves the presence of *both* the thought that P and the thought that Q. On this model, one brings to mind both P and Q, and then owing to the disposition to accept Q upon accepting P, is led to accept Q. So, we might imagine Alice thinking thus: "The market went up by more than 2% today. Wonder if the Fed kept rates steady?" Here she first has both the premise-thought and the conclusion-

thought. Then she thinks "Yes, that's it. The Fed must have kept rates steady."

The problem here is that we are left unable to accommodate cases in which it's plausible that one thinks of and accepts the conclusion at the same time. This seems to be the case when, for example, one thinks and accepts the conclusion Q for the first time, without having antecedently wondered about it. If this is possible, as it seems to me to be, it's implausible to think that one's thinking that Q is part of the *stimulus* condition of the disposition putatively manifested in reasoning.

Moreover it seems that an antecedent disposition of the sort described above is not *necessary* for reasoning. For, in cases where one performs a bit of novel reasoning for the first time, it seems unilluminating to say that one had the disposition to draw the relevant conclusion all along. When Archimedes discovered for the first time the Principle of Buoyancy, is it fruitful to characterize him as then having had all along the disposition to infer the Principle from his evidence about water displacement? If any such disposition was involved here, it would have to involve extremely specific stimulus conditions. Perhaps he had the disposition to come to accept the Principle of Buoyancy if he got into a bathtub at a particular time, when the water had a certain temperature, provided he had woken up when he did, had had a certain type of breakfast, had been worried about disappointing the King, and so on... This is a very specific disposition indeed! And the more specific you make a disposition the less work it can plausibly do in an explanation.

Neither is an antecedent disposition *sufficient* for our purposes. Alice can come to have a disposition (make it as stable as you like) to think 2) upon thinking 1) in ASSOCIATIVE THINKING without its thereby being the case that she *infers* 2) from 1) in episodes when she goes through the thoughts in ASSOCIATIVE THINKING.

3.3 Dispositionalism

3.3.1 The Intuitive Sketch

So what should we add to the claim that an inference from P to Q involves one's thinking that P causing one's thinking that Q? I want to propose the following intuitive idea: in reasoning from P to Q, one must in addition come to have one's confidence in Q be *sensitive* to one's confidence in P. In other words, one must come to have the *disposition* to lower credence in Q when and because one lowers credence in P.[3]

Let me illustrate how Dispositionalism solves the problem we set out with. In ASSOCIATIVE THINKING, Alice's thought 1) may well cause her thought 2). But in going through that series of thoughts, she doesn't thereby acquire the disposition to have her acceptance of 2) be sensitive to her acceptance of 1). That is, she doesn't come to have her views on the aesthetic qualities of London be sensitive to her views on whether Saul Kripke is a philosopher. (This could in principle occur, but then it doesn't seem problematic to say that there is indeed an inference here, however bizarre.) In INFERENCE, things are different. There, she does come to have her confidence in 2) be sensitive to her confidence in 1). So suppose, for example, that she reads 1) in a newspaper and performs the INFERENCE. Then she realizes that the paper she's reading is a year old and thus loses the belief that the market went up by 2% today. We would naturally expect her to then reduce confidence in the proposition that the Fed kept rates steady today.

Several details need to be filled in, but the core idea is that reasoning constitutively involves updating the way in which one's beliefs are sensitive to one another. And this updating must be caused by the thinker's having the "premise-thoughts", or thoughts the contents of which she reasons *from* – rather than, say, a hit on the head, loss of memory,

[3]This describes the simplest case, where the inference from P to Q doesn't occur within a context of supposition, and where one hasn't already made the inference in the past, among other things. I will discuss a general account which handles the more complex cases later on. Also, though this initial formulation deals with credences, we can also put the point in terms of beliefs: in reasoning from P to Q, one acquires the disposition to drop the belief in Q upon dropping the belief in P.

or the like. The way in which our beliefs are sensitive to one another can of course come to change in these latter ways, but those cases do not count as cases of reasoning. An episode of reasoning, then, requires a particular sort of structure of causation.

Before attempting to provide a more detailed form of the intuitive idea suggested here, let me say a few things about what hangs on whether this dispositional view of reasoning succeeds.

First, what we will end up getting is a fully reductive, or "naturalistic," account of reasoning, in the sense that it will not involve *normative* notions such as *reason*, *support*, and so on. Recent work on reasoning has maintained a pessimism as to whether a reductive account can be made to work. For example, on Paul Boghossian's account, "Inferring necessarily involves the thinker *taking* his premises to support his conclusion and drawing his conclusion *because* of that fact."[4] Notice Boghossian's use of the notion of *support*, a normative concept. He concludes the paper saying, "If the present account of reasoning is along the right lines, it opens up the possibility that reasoning poses as much of a challenge to a naturalistic worldview as does consciousness. It makes it difficult to see what naturalistic process inference could consist in."[5] Crispin Wright, who wishes to understand reasoning in terms of the notion of acting or believing for a reason, has echoed this sentiment in a response paper to Boghossian. The notion of acting for a reason, he says, "is a notion that, although it wears causality on its sleeve, has proved resistant to Naturalistic account. The overall orientation in our thinking about inference that I have been promoting in the second half of these remarks therefore promises little to qualify Boghossian's concluding pessimism about the prospects for our achieving a naturalistic understanding of inference."[6] Lastly, John Broome's recent work describes inference in terms of rule-following. However, as Kripke's Wittgenstein argues forcefully, a naturalistic understanding of rule-following may not be forthcoming. In marked contrast with these accounts, Dispositionalism understands reasoning in terms of

[4]Boghossian (2014), p5.
[5]Boghossian (2014), p18.
[6]Wright (2014), p37.

non-normative concepts such as belief, causation, and disposition.[7]

Second, if we are able to give an account of *reasoning* that eschews normative notions, this opens up the possibility that we might be able to give an account of what it is to act or believe *for a reason* in naturalistic terms. For if, plausibly, reasons are considerations that figure in reasoning, and moreover, by reasoning we come to act and believe for reasons, our accounts of reasoning and of what it is to act or believe for reasons must be intimately related.

Lastly, the question of naturalism aside, if Dispositionalism works, we can in effect dodge the worries surrounding rule-following that are invited by accounts such as Broome's. For nothing in the intuitive sketch of Dispositionalism that I gave earlier relies on the idea of a thinker's following a rule.

3.3.2 The Details

Below, I will suggest one way of filling out the details of the intuitive idea explicated earlier, and explore some of the features of the resulting sketch.

I contend that for a thinker to reason, or infer, starting with some set of attitudes $\Psi(\mathbf{P})$ (we can call these "premise-attitudes"), which she has towards a set of claims $\mathbf{P}$, and ending up with attitudes $\Phi(\mathbf{Q})$ (or "conclusion-attitudes"), which she has towards the set of claims $\mathbf{Q}$, is for her to:

1. Occurrently possess the attitudes $\Psi(\mathbf{P})$.

2. Occurrently possess the attitudes $\Phi(\mathbf{Q})$.

[7]It might be wondered whether, when a normative claim occurs within an intensional context, as opposed to on its own, the resulting claim is not perfectly respectable from a naturalistic perspective. The answer is, I think, it depends. Naturalists ought to be able to give a characterization of what it is to have a normative attitude in non-normative terms. If normative attitudes are thought of as *sui generis*, however, then it seems that their existence poses just as much a threat to naturalism as does the existence of *sui generis* normative facts. This worry has been pressed against expressivists about morality. Since one of the primary motivations behind expressivism is naturalism, it has been claimed, expressivists owe us a naturalistic account of what it is to hold moral attitudes – of what it is to think that something is wrong, for example. See Köhler (2013) for an extended discussion of this worry.

3. Update the way in which the holding of the attitudes $\Phi(\mathbf{Q})$ is sensitive to the holding of the attitudes $\Psi(\mathbf{P})$. (In other words, the thinker updates certain of her doxastic dispositions, in the theoretical case.)

AND for (1) to *simultaneously* and *directly* cause (2) and (3) in virtue of the *contents*, **P**.

That is a mouthful. Let me describe some of the conditions stipulated and offer some reasons as to why they might be necessary.

Two preliminary clarifications: Firstly, an inference may occur within a context of supposition, in which case the attitudes $\Psi(\mathbf{P})$ and $\Phi(\mathbf{Q})$ are not actually held, but held only within that context of supposition. Secondly, the account describes a *basic unit* of inference, as it were. Suppose I perform the following chain of inferences: I reason from P_1 to P_2, from P_2 to P_3, and then from P_3 to P_4. In one sense I have reasoned from P_1 to P_4. But in this case the conditions described above won't be satisfied if we substitute 'P_1' for 'P' and 'P_4' for 'Q'. However, this is a feature not a bug.

What are the attitudes $\Psi(\mathbf{P})$ and $\Phi(\mathbf{Q})$? In the case of theoretical reasoning, these are beliefs, or more generally, credences. Since the attitudes are left unspecified in the account itself, we can leave open the possibility of other types of reasoning. In practical reasoning, for example, intentions seem to be involved.[8]

Occurrent vs non-occurrent states: (1) and (2) deal with *occurrent* mental states because reasoning is a process that occurs over a specific period of time. Occurrent beliefs are at the forefront of one's mind, as it were, at the time of occurrence. What are often called "dispositional" beliefs, on the other hand, do not have this feature.[9] Thus, Alice may believe

[8]Perhaps we can also reason to attitudes like wish, hope, and desire. Philosophers have spoken of "rational desires" – see for example Smith (1994).

[9]The term 'dispositional' is a bit misleading here. What is important to keep in mind for our purposes is that the distinction between dispositional and occurrent beliefs in the sense employed here is orthogonal to the distinction between representational and dispositional theories of belief. The latter is a distinction between theories that give an account of what it is for someone to believe something. The former distinction, however, is meant to separate two ways in which a belief – whatever that turns out to be – is present in the mind. For more on these and other relevant distinctions, see Schwitzgebel (2010).

that the Earth revolves around the Sun in the dispositional sense even during INFERENCE, or in the midst of a dreamless sleep. However, she does not *occurrently* believe this proposition during INFERENCE or in the midst of a dreamless sleep. Condition (3), on the other hand, involves straightforward dispositional beliefs or credal states – the sensitivity acquired through the process of reasoning is not itself a temporal event. (Of course, the *acquisition* of the sensitivity is indeed a temporal event.)

Condition (3): This condition says that any inference must conclude with the updating of the way in which a reasoner's credences are sensitive to each other. The sort of sensitivity that results depends on the particular inference and its direction, so to say. I have already described what happens in the case of Alice in INFERENCE, where upon learning some proposition P she comes to believe (or raise credence in) Q. Sometimes, however, we can reason to suspension of belief; we can become agnostic as to whether something is the case. For example, if Alice were to learn that the market news she read was erroneous, she may suspend belief as to whether the Fed kept rates steady. In this case she may come to have, say, 0.5 credence in the proposition that the Fed kept rates steady, but will become disposed to revise that upwards if she learns that the news wasn't erroneous after all.

The causal link between (1) and (3) Why does there have to be a causal link between (1) and (3)? Isn't it sufficient that the disposition(s) mentioned in (3) are acquired, some way or another? There are several reasons why this particular structure of causation is required – one interesting and illuminating case involves forgetting. Suppose P = *I had eggs for breakfast on day D* and Q = *the mailbox key was on the kitchen counter on day D*. Suppose Bob thinks P and then thinks Q, without making the inference from P to Q. Suppose further that Bob acquires the disposition to forget P when he forgets Q. (This is not an artificial example – we often forget chunks of facts together.) If the account did not require a causal link between the thought that P (in virtue of its content) and the acquisition of the disposition to drop Q upon dropping P, it would wrongly entail that Bob reasons from P to Q here. But with the causal condition in place, Dispositionalism can accommodate this

forgetting case – it's not plausible that Bob's thinking that P, in virtue of the content, P, caused him to acquire the disposition to forget P and Q together.

Now, it might be wondered whether condition (1) is the only cause relevant in reasoning. In particular, we might raise the following two questions. Firstly, the nature of one's processes of reasoning can (counterfactually) depend on the sorts of past experiences, education, and the like, that one has had. So aren't these causally relevant to one's processes of reasoning? Secondly, the attitudes described in (1) can only cause (2) and (3) within a particular cognitive system – so isn't the structure of the cognitive system as a whole causally relevant?

These questions are interesting, but apply to causal explanations *in general*. Hence, they pose no particular problem for Dispositionalism *per se*. Suppose for instance that a gas leak causes a fire. There are several events prior to the gas leak without which the fire would not have occurred – the particular way in which the building was constructed, the personal histories of the workers who installed the gas pipes, and so on. Yet we say that the *gas leak* caused the fire. Secondly, the fire could only have occurred if the general conditions were right – only if there was enough oxygen, for example. But the presence of oxygen plausibly figures as a background condition in the causal explanation here, rather than as the cause itself.

Simultaneous and direct causation: Why does (1) have to cause (2) and (3) simultaneously and directly? Because, I think, these conditions are necessary in ruling out bizarre causal processes that satisfy (1), (2), and (3), but nonetheless are not inferential. Thus, suppose, for example, a sophisticated brain surgeon has decided that once you make occurrent some particular beliefs, P and Q, he will implant a doxastic disposition in your brain so that (3) is satisfied. The directness condition rules out this case's counting as reasoning. Consider another scenario. Suppose you think P for the first time, which causes you to think Q for the first time. Now suppose, because P and Q are interesting and controversial propositions, an hour later you are wondering whether there is a connection between them. And then you make the inference: P, therefore Q! Since an inference is a temporally extended process,

we can ask: when did the inference occur? Did it occur right now, or did it start an hour earlier and finish just now? Plausibly, the answer is: it occurred right now. The simultaneity condition in (3) allows us to capture this intuition.

Causation in virtue of contents: In formulating Dispositionalism, I have appealed to the idea of *causation in virtue of content*. This idea has emerged as one of the ways to avoid the problem of deviant casual chains *à la* Davidson. The thought is this: we can talk not only of events causing other events, but also of event e_1 causing e_2 in virtue of property p. Thus suppose the throwing of a brick at a window caused the glass to break. We can ask, in virtue of what did this event cause the glass to break? Plausibly, in virtue of the brick's momentum as opposed to its color. Similarly, we can talk about one thought causing another in virtue of its content as opposed to its form, its occurring at a particular time, etc.[10]

We now have the resources to deal with the following case, which involves beliefs about one's own beliefs. Suppose Alice has recently been thinking about Moore's Paradox. She looks out the window, noticing that the streets are wet. She then forms the belief, P: *It's raining.* This causes her to think Q: *I believe it's raining.* If Alice is like the rest of us, she comes to have the disposition to drop Q upon dropping P. Yet, in this case she need not have reasoned from P to Q. What I want to claim here is that it's not accurate to say that her thinking Q and her acquiring the aforementioned disposition are caused simply by her thinking P in virtue of the content, P – there is more to the causal story. The associative leap from *P* to *I think that P* would have occurred for other values of P, and this is plausibly owing to a (fleeting) disposition she formed after thinking for a while about Moore's Paradox. So at the very least, I think, this disposition figures in the causal story. And *this* is why the episode does not constitute reasoning.

There is doubtless much more to be said here. But as a general point, I think there are other cases where an additional causal influence can affect whether a particular process is

[10]For more on how the idea of causation in virtue of contents may obviate the problem of deviant causal chains, see Schlosser (2007). To see the idea employed to understand the phenomenon of acting for a reason, see Arpaly and Schroeder (2013); Wedgwood (2006) applies the idea to give an account of reasoning.

of a particular kind – so *that* feature of the account shouldn't be a problem. Suppose, for instance, the kick of a soccer ball causes it to go into the back of the net. Is this event a goal? It depends. If the kick caused the ball to go there by itself (assuming there was no offside, etc.), then the event is a goal. But if the kick, followed by the punching of the ball by a member of the same team, *jointly* cause the ball to go to the back of the net, then the event is not a goal. Myriad other examples, including several from the natural sciences, can be found as well.

Reasoning within contexts of supposition: Not all cases of reasoning involve the acquisition of new conclusion-beliefs. For example, a detective may reason in the following way: "Suppose the butler didn't commit the murder. Then the gardener did – he is the only other person without an alibi." In reasoning thus, the detective does not come to have the belief that the gardener committed the crime (though perhaps she acquires the conditional belief: if the butler didn't do it, then the gardener did). Rather, the detective reasons within a context of supposition. Let $\Psi(\mathrm{P})$ be the acceptance of the proposition that the gardener doesn't have an alibi, let $\Phi(\mathrm{Q})$ be the acceptance of the proposition that the gardener committed the murder, and let the context of supposition be one where it's assumed that the butler didn't commit the murder. We can now understand the detective's reasoning in the following way: within this context of supposition, the detective comes to have her acceptance of Q be sensitive to her acceptance of P – for, if she learns that the gardener does have an alibi, she will lose her acceptance of Q, within the context.

Condition (3) and confirmatory reasoning: *Prima facie*, condition (3) seems to rule out cases of confirmatory reasoning, in which one performs an inference that confirms what one already believes and why. Suppose the detective thinks to herself "Why think the butler committed the murder, rather than the gardener, again? Oh right, because the butler is in the will." In this scenario, the detective repeats a bit of reasoning that she has already performed, and thus acquires no new beliefs (not even a conditional belief). In repeating this bit of reasoning, I claim, she nevertheless updates a doxastic disposition she has. This can

happen in two ways. One possibility is that she makes her confidence in the claim that the butler committed the crime *more sensitive to* her confidence in the claim that the butler is in the will – in other words, she becomes disposed to reduce to a greater extent her credence in the proposition that the butler committed the crime upon learning that the butler is not in the will. The other possibility is that she makes her disposition more robust, or resilient. In general, sensitivity amongst the various credences we have can fade, and sometimes we need to redo some reasoning to bolster our doxastic dispositions – sometimes we need to reassure ourselves.

But *must* we reason by updating our doxastic dispositions in this way? What happens when one goes through a pattern of reasoning from P to Q where one is already sure of Q, and one's belief in Q is supported by one's belief in P? What happens, for example, when a mathematician teaches a proof that he knows by heart to his students? I contend that there is a difference between genuine reasoning and the mere *recital* of a bit of reasoning – just as there is a difference between genuine mental multiplication and the mere recital of a times-table. Thus our imagined detective may go through her reasoning as to who the murderer is with her partner without thereby updating her relevant doxastic dispositions. She already knows why the butler is most likely the culprit, but is merely trying now to convince her partner. But here I think she is simply reciting a bit of reasoning that she has already performed – without the updating of doxastic dispositions, no genuine (theoretical) reasoning can occur.

3.3.3 Dispositions and Masking

Sometimes, we are already fairly certain of some proposition. Can we then reason to it? Suppose I'm already fairly certain of Q, because of consideration R. Does this mean that I cannot reason to Q from P? On the face of it, Dispositionalism would seem to rule out such reasoning. For suppose I "reason" from P to Q. Because my belief that Q is already sustained by my believing R, it seems that my lowering confidence in P would not lead to

my lowering confidence in Q. Therefore Dispositionalism would have to maintain that an inference from P to Q is not possible in this case. But this seems wrong.

Thus suppose I see the car in the driveway and reason that someone is home. I daydream for a while as I walk closer to the house. When I am near the front door, I see a light on in the kitchen area, and then I reason again – this time from a different consideration, namely, that a light is on in the kitchen area – that someone is home. This seems possible, yet its possibility doesn't seem to depend on its being the case that *were* I to lose the belief that a light is on (maybe it was coming through the neighbor's window after all) I would then reduce confidence in the claim that someone is home.

I contend that Dispositionalism *can* accommodate such cases. But its doing so, given the way it has been presented above, relies on our not understanding dispositions in purely counterfactual terms. We must, that is, reject an analysis of dispositions in counterfactual terms. So to say that a thing T is disposed, when given a stimulus S to produce effect E is not simply to say that T *would* produce E *were* the stimulus condition S to obtain. Dispositions can be *masked*.

Thus consider a light source L that is connected to an electric outlet E_1. L has the following disposition: to turn off upon being disconnected from E_1. Now suppose we connect L to an independent power source E_2, which is able to sustain L on its own. Has L now lost its original disposition to turn off upon being disconnected from E_1? We can say: no, the disposition has not been destroyed, but is merely masked by the connection of L to E_2. I think this is a natural way of understanding the dispositions involved.

Notice that this example is exactly analogous to our case of beliefs with multiple sources of support. So, when I see the light and reason that someone is home, I *do* gain the disposition to reduce confidence in the proposition that someone is home upon reducing confidence that the light is on. It's just that this disposition is *masked* by the way my belief that the car is in the driveway sustains my belief that someone is home.

Even if we insist on understanding dispositions in counterfactual terms, we need not give

up on Dispositionalism. In that case, the dispositions involved will have to be complicated. So when one already believes Q because of R, and then reasons from P to Q one gains the disposition to: reduce confidence in Q upon (sufficiently) reducing confidence in R, or losing the belief that R, or rejecting the reasoning from R to Q, or... AND reducing confidence in P. However, since counterfactual analyses of dispositions have been unpopular for independent considerations involving finkish dispositions, there is no reason to build in such complications into Dispositionalism.[11]

3.3.4 Impossible Inferences and Logical Entailment

Can a thinker infer P∨¬P from P? Bizarre cases aside, one's credence in P∨¬P will not be sensitive to one's credence in P. So, Dispositionalism is committed to saying that this is impossible. And this might seem to be a problem with Dispositionalism.

I don't think it is. The tendency to see it as a problem comes from conflating rules of logic, or entailment, with patterns of reasoning.[12] I claim that we do not, and cannot, infer P∨¬P from P, except in bizarre cases.[13] Of course, we might use the fact that P entails P∨¬P in a logical proof. But here we are following a rule (of implication) within a certain context; merely following a rule does not guarantee that one is reasoning. I can decide to follow this rule: Do not step on any cracks on the sidewalk. In following this rule, I'm not *ipso facto* reasoning in the sense we are interested in (of course, this is not to deny that there is cognitive processing involved in my following this rule). This point will be important in what follows.

[11] See Lewis (1997) for a discussion of finkish dispositions and reasons to think that counterfactual analyses of dispositions cannot be true. Lewis understands the "simple conditional analysis" in the following terms: "Something x is disposed at time t to give response r to stimulus s iff, if x were to undergo stimulus s at time t, x would give response r" (p133).

[12] See Harman (1986) for a detailed defense of this claim.

[13] I leave the qualification here for the following reason. Perhaps we can imagine a bizarre case where a thinker comes to have her credence in P∨¬P be sensitive to her credence in P in the way Dispositionalism requires. I'm fine with saying that such a thinker reasons from P to P∨¬P.

3.4 Broome's Rule-Following Proposal

John Broome has recently developed a rule-following account of reasoning. On his view, when you reason, you consciously operate on the premises (the contents, or propositions, not sentences) to construct to conclusion, by following a rule. So, to use Broome's example, when you reason from P and if P then Q, to Q, you follow the rule of Modus Ponens. However, in applying a rule you need not know explicitly what the rule is – just as, in applying rules of grammar, we often do not know explicitly the contents of the rules we apply.

Further, Broome claims, your performing a bit of reasoning entails that you have what he calls a First Order Linking Belief. In the Modus Ponens case, the belief will be: P and if P then Q entail Q. Importantly, the First Order Linking Belief *cannot* be a *premise* in the reasoning. Else, we are led into the famous Lewis Carroll regress – if we add if P and if P then Q, then Q as a premise, then we will need another linking belief to get to the transition to Q, at which point we will have to add if P and if P and if P then Q, then Q, and so on.[14] So: the Linking Belief cannot be a premise. But how do we distinguish what is a premise in reasoning from what is not in this case? Broome's answer is that the Linking Belief is not conscious, whereas the premises are all conscious. Besides, since Broome is interested in giving an account of conscious, active reasoning, it's important for his view that the premises be consciously entertained.

Broome insists that reasoning is *active* – it cannot just happen to you. Moreover, he thinks, the rule-following nature of reasoning explains how reasoning can be an *act* of ours. Broome says, "Reasoning is an act because in reasoning you follow a rule. The rule does not merely cause you to behave in a particular way, as a program does to a computer. The rule *guides* you and you actively follow it."[15]

Now, what differentiates following a rule, and thereby reasoning, from merely thinking

[14]See Carroll (1895).
[15]Broome (2013), p236.

associatively? What is it to be "guided by" a rule? Broome's thought is that when you follow a rule, the process *seems right* to you. Importantly for Broome, 'seeming right' does not refer to a phenomenal state. Neither is seeming right a kind of endorsement of a particular rule; for Broome, applying a rule in reasoning automatically counts as an endorsement of the rule. Rather, seeming right involves having the *disposition to be open to correction*. Broome writes:

> When a process seems right to you, you are open to the possibility that the process might no longer seem right to you if a certain sort of event were to occur. We may call the event 'checking'.
>
> Checking may consist simply in a repetition of the process, or it may consist in a different process. Here is an example. If you are asked 'Three fours?', you will probably spontaneously answer 'Twelve', and this will seem right to you. You may later check your conclusion by calling up a spontaneous response once again, or alternatively you may do the sum by counting on your fingers.
>
> Your openness to correction is a disposition. You are disposed to lose the attitude of seeming right in particular circumstances – specifically if checking occurs and produces a different result.[16]

3.4.1 The Problem

The chief problem with Broome's account, however, is that one can decide to follow a rule of associative thinking, as it were. Here is an example. I can decide to follow the bizarre rule that whenever I think to myself "Saul Kripke is a philosopher" I shall immediately think "London is pretty." It's hard to see what would stop me from being able to do so in principle. Surely, following such a rule has to be *possible*. And surely, this need not entail that I am reasoning when I follow this rule.

Furthermore, my following this rule can be something I *do*, just as much as any other putative mental act. Moreover, I can follow the rule correctly or incorrectly. For, if I entertain the thought that Saul Kripke is a philosopher, and then go on to think immediately afterwards that Saul Kripke has a beard, I have followed the rule incorrectly; thus my doing so can "seem wrong" to me relative to the rule. So I can be open to correction in this way.

[16] Broome (2013), p238.

Yet, all this is associative thinking; or if you like, it's just thoughts following thoughts in a non-inferential way. If I follow the aforementioned rule, I am not, I hope, *ipso facto* performing an inference. Thus Broome's view, I think, ultimately lacks the resources to separate cases of reasoning from cases where one thought follows another in a non-inferential way.

Dispositionalism, on the other hand, can give us the verdict that I am not reasoning when I follow the bizarre rule above. For, when I follow that rule, I do not thereby gain the disposition to reduce confidence in the proposition that London is pretty upon reducing confidence in the proposition that Saul Kripke is a philosopher (if I do then I am reasoning, albeit in a very bizarre way!). My losing the belief that Saul Kripke is a philosopher will not affect my aesthetic evaluation of London. *That's* why I am not reasoning.

What if I decide to follow the rule that my credence in the proposition that London is pretty is to co-travel with my credence in the proposition that Saul Kripke is a philosopher? I'm not sure that this is possible – I'm inclined to think that we can't just decide what to believe (except in an indirect way). I cannot get myself to believe that the Earth is a pentagon even if I try. But if we waive such worries for the moment and suppose that it is possible to adopt this rule, then I'm fine with saying that in adopting this rule, one has adopted a rule of *inference* as opposed to a rule of associative thinking. The key point is that Dispositionalism can differentiate in this way between rules of inference and rules of associative thinking. What Broome's account lacks the resources to do, I claim, is to allow for the fact that one *can* adopt a rule of associative thinking – and this is not a drawback that is shared by Dispositionalism.

3.4.2 Possible Responses

Here are a few possible points that could be raised in response to the worry sketched above. I think that none of them succeed in deflecting the worry.

Possible Response # 1: In the bizarre example above, I do not "operate" on the content

that Saul Kripke is a philosopher in the proper way. Or alternatively, the way in which it "seems wrong" to me to think that Saul Kripke has a beard, after thinking that Saul Kripke is a philosopher, is not the sort of seeming wrong that is involved in reasoning.

However, this is entirely unsatisfactory as a response to the challenge because what we're trying to find out is precisely what it takes for someone to operate on contents and for something to seem wrong in the way required for reasoning.

Can the idea of *causation in virtue of content*, which I have appealed to in formulating Dispositionalism be of help to Broome's account here? Plausibly, no – for, my thinking "London is pretty" seems to be caused by my thinking "Saul Kripke is a philosopher" in virtue of the latter thought's content.[17]

Possible Response # 2: In order to reason from P to Q, as Broome writes, there must be "a semantic connection between the attitudes' contents."[18] That's why the bizarre mental process in the example doesn't count as reasoning.

But surely I can reason in the following way: "There is a blue Yaris in the driveway. So, Bill must be over for dinner." While there may well be an *evidential* connection between these claims, there is no *semantic* connection. One way to ensure a semantic connection is to require that I must consciously think "if there is a blue Yaris in the driveway, Bill must be over for dinner" in making the inference. But this is phenomenologically implausible.

Possible Response #3: The bizarre example above does not entail that I have the First Order Linking Belief: if Saul Kripke is a philosopher, then London is pretty.

The plausibility of this response depends on what it takes for one to have the linking belief. Broome says "we may treat your disposition to derive the conclusion from the premises – where deriving it includes actively operating on the premises following a rule and coming to believe the conclusion – as itself constituting an unconscious, implicit belief the premises

[17]For more discussion on how associative thinking can involve causation in virtue of content, see Arpaly and Schroeder (2013), ch3.

[18]Broome (2013), p231.

imply the conclusion."[19] If that's right, then the linking belief is not so much an independent necessary condition for some process to be reasoning, but rather is, in Broome's own words, "entailed by the condition that you come to believe the conclusion by operating on the premises in the way I have described."[20] But if that's true, then unless we're given an account of what it takes to "operate on the premises" that precludes the bizarre example from counting as reasoning, then there is no reason not to attribute to me the "linking belief" that if Saul Kripke is a philosopher, then London is pretty.

Broome's account would thus yield that it's *impossible* for me to decide to follow a rule of association, so that I merely decide to think one thought followed by another. But if we think this is possible, then we should reject Broome's rule-following proposal.

3.5 Causation in Virtue of Rationalization

I said earlier that no relationship between the relevant contents themselves will suffice to render a process reasoning. A family of views takes this point on board but nonetheless reserves an essential causal role for a certain kind of relation between contents. I have in mind here Arpaly and Schroeder (2013) and Wedgwood (2006).

The core idea is this. When someone reasons from P to Q, her coming to accept Q is at least partially explained by the *rational connection* between P and Q. What exactly the rational connection comes to will depend on what the best first-order theory of practical and epistemic normativity turns out to be.

Arpaly and Schroeder write:

> *To think or act for a reason* is for the event of one's thinking or acting to be caused (or appropriately causally explained) by one's other attitudes in virtue of the fact that these attitudes rationalize (to some extent) the thought or action.[21]

[19] Broome (2013), pp233-34.
[20] Broome (2013), p234.
[21] Arpaly and Schroeder (2013), p62.

Let me illustrate with an example. Suppose for a moment that the correct theory of epistemic rationality turns out to be the following: a claim P supports a claim Q just in case P logically entails Q. Then, for it to be the case that I reasoned from some P to some Q, my coming to accept Q must be caused by my thinking P, *in virtue of the fact that* P entails Q.

While Wedgewood doesn't tell us how we should accommodate cases of *bad* reasoning, Arpaly and Schroeder do allow for some such cases. This is the function of the hedging qualification "to some extent" in their formulation above. So what happens when Bob goes out to smoke a cigarette because it will be pleasurable, despite knowing that smoking can cause cancer? He still acts for reasons because his belief that smoking will cause such-and-such a pleasurable sensation (along with, if you like, the desire to have that sensation) does rationalize his going out to smoke. Although his beliefs (and desires) as a whole do not support his smoking (it can cause cancer!), his smoking is to some extent rationalized by his belief (and the relevant desire) that it will lead to such-and-such a pleasurable sensation.

3.5.1 The Problem

I think there is a powerful objection to the Arpaly-Schroeder view, however. The problem is that one can reason to Q from P even if P is *no reason* to form the belief or intention Q, even relative to one's other attitudes.

Let me give an example. Bill lives in a modern society, is not a hermit, is reasonably well-read, up to date on the news, etc. Now, Bill knows that Bob is a homosexual. He takes this to be a reason to not hang out with Bob. So consider this bit of practical reasoning that Bill might engage in:

1. Bob is going to be at the Party.
2. Bob is homosexual.
3. I will not go to the Party.

Here he forms the intention not to go to the Party because Bob will be there and Bob is homosexual.

But, plausibly, Bob's being homosexual is *no reason* at all for Bill not to hang out with him. And given that Bill lives in modern society, he has ample evidence that supports this claim. So when Bill forms the intention not to go to the Party, his intention is to *no extent* rationalized by his other attitudes (namely that Bob will be at the Party and Bob is homosexual) *even relative* to Bill's total set of beliefs. However, surely Bill can reason in this way. Thus, the Arpaly-Schroeder view is false.

Now perhaps Bill has a background belief to the effect that homosexuality is bad. However, I want to contend that the other evidence that Bill has undermines the justification for this belief – so he shouldn't have this belief in the first place. And since Bill's other evidence undermines the justification for this belief, this belief cannot make it the case that (1) and (2) rationalize (3).

The best way to avoid this objection, I think, is to adopt the Humean Theory of Reasons. On this view, any desire whatsoever to some extent rationalizes an action that would satisfy the desire. So if Bill has a desire not to hang out with Bob because Bob is homosexual, this to some extent rationalizes Bill's not going to the Party (even though the balance of reasons may weigh in favor of going to the Party.)[22] However this means that the defender of the Arpaly-Schroeder view is committed to accepting the Humean Theory of Reasons, and thus to a specific metaethical thesis that is very much up for debate.

On the contrary, Dispositionalism does not carry such metaethical commitments and thus, I believe, wins out in this respect. For according to Dispositionalism, it need not be the case that the premises from which one reasons rationalize (to any extent) the conclusion that is drawn. All that is needed is that the conditions (1), (2), and (3) are satisfied as described in the formulation of the theory. Dispositionalism does not entail, then, that Bill

[22]See Schroeder (2007) for a recent defense of this view. Schroeder provides an extensive discussion of what he calls the "Too Many Reasons" problem, which is relevant here. In the case I've discussed, the Schroeder-style response would be this. The fact that Bob is a homosexual is indeed a reason for Bill to avoid the Party. However, because the reason is very weak, it is pragmatically bad to say that Bill has this reason. The key move that Schroeder makes here, that would allow him to say this, is that the strength of a desire can come apart from the weight of the reason that corresponds to it. For, Bill's desire to avoid Bob because Bob is homosexual may be fairly strong, and we don't want to say that this gives rise to a strong reason to avoid Bob – this would be implausible on first-order ethical grounds.

cannot reason in the way described above.

3.6 The Doxastic Proposal

One way to avoid the kind of objection discussed above is to claim the following. When a thinker reasons from P to Q, (to use the simplest case in which the reasoner ends up believing both P and Q) she *takes* P to *support* Q. Furthermore, her taking P to support Q is a mental state that plays the appropriate causal role in getting her to transition to the belief that Q. However, it's irrelevant whether P *actually* supports Q.[23]

Now what is one's "taking P to support Q"? For all we've said, it could be one of several things. Maybe it's a particular belief; maybe it's the endorsement of a rule of inference; perhaps it's some other, non-cognitive state of mind.

According to one view discussed in the recent literature, one's taking P to support Q amounts to a *doxastic state*, i.e. a belief. Let's call this view the Doxastic Proposal. Thus the proposal is one way of cashing out Frege's thought that "To make a judgment because we are cognizant of other truths as providing a justification for it is known as *inferring*"[24] or Boghossian's related proposal that "S's inferring from p to q is for S to judge q *because* S *takes* the (presumed) truth of p to provide support for q."[25]

Note that for the Doxastic Proposal to be plausible, the support relation must be seen as holding between the propositions in question, namely P and Q, and *not* one's attitudes towards these propositions. In other words, it's not plausible that when you reason from P to Q you must take your *believing* P to support your *believing* Q. (Your believing P is no epistemic reason, except in some rare cases, for you to believe Q!) Rather, what is plausible, is that you take (the presumed truth of) P itself to support Q. So, Philip Pettit, a defender of the Doxastic Proposal, claims that what is involved in an inference is a "meta-

[23]This is the view defended in Boghossian (2014).
[24]Frege (1979), p3.
[25]Boghossian (2014), p4.

propositional" attitude, by which properties like truth, evidential support, inconsistency, and so on, are ascribed to propositions themselves, as opposed to our own attitudes towards propositions.[26]

3.6.1 The Over-Intellectualization Problem

A major issue with the Doxastic Proposal is that it runs the risk of over-intellectualizing the process of reasoning. That is, it makes the process of reasoning too intellectually sophisticated for certain agents who nonetheless can plausibly reason. Here is Boghossian's succinct description of the worry:

> A child, we are inclined to think, can reason. Luke and Drew are playing hide-and-seek. Seeing Drew's bicycle leaning against the tree, Luke thinks: "If he were hiding behind that tree, he would not have left his bicycle there. So, he must be behind the hedge." That looks like reasoning. But do children have meta-beliefs about the relations between their premise judgments and their conclusions? Do children have the concepts of premises and conclusions? Do they have the normative concept of one belief justifying another? ... Do Luke and Drew, playing hide-and-seek, really need to have the concept of a proposition, and the concept of one proposition following from another?[27]

I'm quite inclined to agree with Boghossian here. If we require that reasoners have, and employ in the right way, the concepts of evidential, or practical, support, we will classify many children as not being able to reason, when intuitively they can.

But concept attribution can get murky. A defender of the Doxastic Proposal might say: indeed, Luke and Drew have the requisite concepts – they just can't verbalize them yet. This would lead to a dialectical impasse. So let me assume that the over-intellectualization problem can be solved, or waived. Below, I will raise a dilemma for the Doxastic Proposal that seems to me to constitute a more decisive objection.

[26] See Pettit (2007), p233-238.
[27] Boghossian (2014), pp 6-7.

3.6.2 A Dilemma for the Doxastic Proposal

In reasoning from P to Q, the Doxastic Proposal says, a thinker must believe that P supports Q, and this belief must play a certain causal role in getting her to conclude Q. Let's focus on the belief that P supports Q. Is it conscious or unconscious in the reasoning process?

Suppose it's conscious. Now we run into two major problems. Firstly, it's simply not phenomenologically accurate to characterize reasoners as consciously entertaining some belief about support *every time* they reason. In INFERENCE, Alice can go directly from thinking 1) to thinking 2) without any mediation of conscious beliefs about support. If we required such beliefs as essential components of reasoning, it would follow that we hardly ever reason at all.

Secondly, we are now left without the resources to explain why the belief that P supports Q is *not* a premise, and thus are left without resources to evade the Lewis Carroll regress. Here is Boghossian explaining the problem:

> But what exactly is the difference between a premise belief and a belief that plays a role in the background of an inference? Presumably, it is the difference between a belief on which the conclusion is *based* versus a belief that, while somehow involved in the inference, is *not* one on which the conclusion is based. But the notion of a conclusion's being based on a premise judgment is precisely the notion we are trying to understand, in investigating the nature of inference. So in employing the notion of a belief that plays a role as a mere background, and so not one on which the conclusion is based, we seem to have helped ourselves to the very notion at issue.[28]

Boghossian here takes this to be a problem for the Doxastic Proposal in general. However, it is not. Recall that Broome had a useful way of distinguishing premise attitudes from background attitudes in reasoning – the former, but not the latter, are *conscious*. So, the defender of the Doxastic Proposal can respond: the belief that P supports Q is not a premise because it is *unconscious*.

[28] Boghossian (2014), p8.

This brings us to the second horn of the dilemma. Let's suppose that the belief is unconscious. But if we think this, we should also think that it's enough if the belief is implicit – i.e. present, even if never entertained before. (For example, you have the implicit belief that the number of planets in the Solar System is less than 342, even though, quite probably, you never explicitly had this thought before.) It would be an odd requirement – problematic for the reasons mentioned in the first horn – to hold that for an inference to be possible the support-belief must have been entertained before. Alice, when she performs INFERENCE need not have entertained the belief that 1) supports 2) before. For suppose Alice is performing INFERENCE for the *first time*. If we require that she has the conscious belief that 1) supports 2) in performing this first-time inference, we run into the same two problems rehearsed in the first horn.

So the unconscious belief that P supports Q can be *implicit*. This way we can say it plays the role of a *background condition* rather than a premise. In this way we avoid both the problems of the first horn. Phenomenological implausibility vanishes because we need not be aware of our implicit beliefs, and the Lewis Carroll problem is avoided because we can say: since the belief is unconscious it's not a premise-attitude.

But now what makes it the case that you have the implicit belief that P supports Q, whenever you reason from P to Q? We can't say: it's in virtue of your inferring (or having the disposition to infer) Q from P that you implicitly believe that P supports Q. This is viciously circular as an account of inference. But what else can we say?

We can't say that it is in virtue of your being disposed to *state* P as your reason for believing Q that you have this implicit belief either. That can't be what *makes it the case* that you have the belief. As Boghossian argues compellingly:

> the property of a person's thinking something *for a reason* is not response-dependent. To say that R was S's reason for A'ing implies that S took R to support his A'ing at the time that he A'ed, and that his so taking it led to his A'ing.
>
> I don't see how S's being disposed to *say* that R was his reason for A'ing could

> *make it the case* that he took R to support his A'ing and that this taking had a certain causal impact. His saying it might be very good evidence that R was his reason. But saying that R was his reason can't be *constitutive* of R's being his reason.[29]

The only options left, it seems to me, are the following two:

1. We can take the notion of an implicit background belief as *sui generis*, so that while such beliefs play an essential role in reasoning, they cannot be helpfully cashed out in other terms. This proposal is unsatisfactory because what we are looking for is an *explanation* of the process of reasoning. And if that's the goal, it's unhelpful to simply posit an entity – in this case a *sui generis* implicit belief – that is stipulated to do all the necessary explanatory work. A better account would tell us what such an implicit belief consists in, how it relates to other mental entities we're familiar with, how it brings about the thinker's reaching a conclusion, and so on.

2. The other option is to cash out the idea of an implicit background belief in dispositional terms which don't presuppose the notion of inference. But now, this project is no different from the project of Dispositionalism as described earlier. The difference becomes merely verbal – whereas I have resisted using the term 'implicit background belief' to describe the dispositions involved in reasoning, the theorist we are imagining here insists on using that terminology. So if the Doxastic Proposal is interpreted along these lines it collapses, for all practical purposes, into a dispositional account of reasoning.

3.7 Rejecting the Taking Condition

What Boghossian calls the "Taking Condition" is a natural and intuitive characterization of what is involved in inference. What's the difference between merely thinking P and then thinking Q as opposed to *inferring* from P to Q? The thought goes: in order to make the inference, one must take P to support Q and then conclude Q because of this recognition.

[29] Boghossian (2014), pp 10-11.

The idea is so natural that Frege and Boghossian, among others, find it obvious – any account of inference, on their view, must incorporate the taking condition in some way. Boghossian writes, "Any adequate account of inference, I believe, must, somehow or other, accommodate [the taking] condition...The intuition behind the Taking Condition is that no causal process counts as inference, unless it consists in an attempt to arrive at a belief by figuring out what, in some suitably broad sense, is supported by other things one believes."[30]

However, if what I have argued so far is right, this picture must be rejected. I have argued that the Doxastic Proposal must collapse into a dispositional account of reasoning. Furthermore, I argued earlier that the Antecedent Disposition Thesis is false. That is, the dispositions that are essential to reasoning are acquired through the process of reasoning rather than being a precondition for reasoning.

If this is right, then there is no mental state, or set of states, for the reasoner's "taking P to support Q" to correspond to, so that this taking is prior to the inference, as Frege and Boghossian have in mind. Of course the set of dispositions acquired through a process of reasoning could indeed be fruitfully characterized as one's taking the premises to support the conclusion. But then one's taking the premises to support the conclusion turns out to be the *product* rather than the *precondition* of reasoning.

3.8 Practical Reasoning

So far I have focused on theoretical reasoning, that is, reasoning that involves only beliefs or credences, chiefly because some of the issues in giving an account of inference are best brought out in the theoretical case. But now that we have an account of theoretical inference in hand, we can easily generalize this to practical reasoning, and perhaps other kinds of reasoning, if they exist.

Practical reasoning, it seems to me, concludes in *intentions* or more generally, *plans*. The

[30]Boghossian (2014), p5.

premise-attitudes may involve intentions, but need not.

Consider an example. Suppose Alice is deciding where to go for lunch. She needs to be back at the conference by 2 pm, and it's half-past noon now. So she has time but not too much time. Recalling that Minca's ramen is really tasty and freshly prepared, and the line is often not bad, she decides to go there. This concludes her practical reasoning – she ends up with a plan. If Dispositionalism is right, then her conclusion – the plan to go to Minca for lunch – will be sensitive to her continuing to hold the premise-attitudes, which in this case are beliefs to the effect that 1) Minca serves good ramen and 2) the line is generally not bad. And this seems correct as a description here. For suppose Alice goes to Minca and is told that the wait is an hour long. We'd then expect her to go elsewhere (unless she performs a new line of reasoning and decides to be late for the next talk).

Our practical reasoning can occur in suppositional contexts as well – in this way we come to have contingency plans. For example, Alice might want to get coffee after lunch but might not want to be late for the next talk. So she might decide: if I'm done with lunch with twenty minutes to spare, I'll look for a nearby coffee shop. We can understand her reasoning as concluding in the decision to look for a nearby coffee shop under the context of supposition wherein she finishes lunch with twenty minutes to spare.

3.8.1 Robustness of Intention

One key apparent difference in the practical case, however, is that there doesn't seem to be a notion of "degree of intention" corresponding to the notion of degree of belief, or credence, that we made use of the the theoretical case. Thus suppose Alice, on her way to Minca, realizes that there is a parade blocking one of the streets. This will mean a delay of at least ten minutes or so. And this is a reason against going to Minca – at least in *this* respect, a restaurant that is on Alice's side of the parade route is preferable. But Alice nonetheless decides to go to Minca, even if she registers the aforementioned reason not to go to Minca. What has happened? In an analogous case having to do with theoretical reasoning we could

just say that Alice reduces her credence in the proposition in question. What should we say in this practical case? Has she reduced the "degree of her intention" to go to Minca?

Though I think that the idea of a degree of intention would be hard to conceptualize, there is a sense in which some intentions are more *robust* than others. My intention to write this paper, for instance, is more robust than my decision to have bean soup for dinner this evening. The possible worlds, if you like, where I abandon my intention to finish this paper are much farther away from the actual world than the possible worlds where I decide to have something else for dinner.

What's going on with Alice above, then, is that the presence of the parade makes her intention to go to Minca less robust. She is more likely to decide not to go to Minca, now that the parade is in her way. For instance if Alice were now to get a call that she really needs to answer, she might simply abandon the intention to go to Minca. However, she might not have done so if she'd only received the call, but the parade wasn't taking place.

We can thus explicate the simplest case of practical reasoning in the following general terms. When an agent reasons from P to an intention to ϕ, she updates the way in which the *robustness* of her intention to ϕ is sensitive to her belief (or credence) in P. The limiting case of reducing the robustness of an intention is simply dropping the intention.

Put in these general terms, the formulation in terms of robustness of intention helps to make sense of cases where an agent comes to intend to ϕ for more than one reason. Suppose A reasons from P to the intention to ϕ. Suppose later on she also reasons from Q to the intention to ϕ. This latter inference, supposing P and Q are independent reasons to ϕ, will make the intention to ϕ more robust. And if A realizes, for instance, that Q is not true, then we'd expect her intention to ϕ to become less robust.

3.8.2 Other Kinds of Reasoning?

The possibility that there could be other sorts of reasoning, apart from theoretical and practical, has received almost no attention in the recent literature on reasoning. Let me

briefly discuss a case which makes it *prima facie* plausible that there may be other kinds of reasoning.

Alice believes that the way to create jobs and to increase the median standard of living in her state is for the state government to adopt a certain Fiscal Policy. She believes that Bob the candidate will push for Fiscal Policy, if elected. She thus comes to hope that Bob wins the election.

In this case, the concluding attitude seems to be a hope, or wish. She need not believe that Bob will win – in fact she may think it's likely that he will lose. Furthermore, it's not necessary that her hope could be reduced to a conditional intention – namely to put Bob into power if she had the means to do so. For, Alice may believe in the sanctity of democracy – so she would not look to simply put Bob into power if the majority did not vote for him, even if she were omnipotent.

Moreover, it's implausible to say in this case that her hope that Bob will win is an irrational, or arational, urge – like the urge to eat chocolate cake or smoke a cigarette. Rather, her coming to hope that Bob will win looks structurally similar to the cases of practical and theoretical reasoning discussed earlier. There seems to be good reason, then, to think that we can reason to a hope, and that such reasoning is not simply an instance of theoretical or practical reasoning.

3.9 Conclusion

In this chapter, I have defended a reductive account of reasoning – Dispositionalism – which analyzes the process in terms of the key concepts of disposition and causation, *inter alia*. I have done so in part by showing how Dispositionalism is able to separate associative thinking from genuine reasoning. Also, I have argued that the other accounts discussed in the literature are either unable to accomplish this task or fail for other reasons.

Let me conclude by sketching how we can use a reductive account of reasoning to give

an account of what it is to believe or act *for reasons*. Start with a truism: reasons are those things that play the "reason-role" in reasoning, or deliberation. In other words, reasons are premises in reasoning.[31] Take the case of Alice in INFERENCE. It seems there that she believes 2) for the reason that 1). That the market went up by 2% – let it be a proposition or state of affairs as you like – plays the reason-role for her, in that she believes 2) for this reason.

What does it take for 1) to play this reason-role? Well, according to Dispositionalism, Alice's thinking 1) needs to cause (perhaps jointly, with her apprehension of other reasons) her thinking 2) and it needs to cause her to acquire a certain disposition – namely, to reduce credence in 2) upon reducing credence in 1). If these conditions are satisfied, then Alice believes 2) for the reason that 1). More strongly, *what it is* for Alice to believe 2) for the reason that 1) is for these conditions to be satisfied.

The analysis is admittedly very schematic, and I don't have the space to discuss it further here. But what I want to suggest is that we should look more closely at the prospect of understanding believing or acting for reasons in terms of reasoning, rather than the other way around. An account of reasoning, then, might have broader implications than hitherto acknowledged for our understanding of reasons and rationality.

[31] See Setiya (2014) for a recent, detailed explication and defense of this idea.

Chapter 4

Evidence and Obligation

4.1 Introduction

Uncertaintists have argued that what an agent subjectively ought to do is partly determined by her moral beliefs. Externalists have argued that an agent's moral beliefs are irrelevant in determining subjective obligation. This chapter defends the claim that an agent's beliefs, be they moral or non-moral, are irrelevant in general in this regard. Rather, an agent's subjective obligations are determined by her *total evidence*, along with her circumstances. The chapter ends by sketching a theory as to how an agent's evidence determines her subjective obligations. The notion of human virtue forms an essential component of the theory, and in this sense, virtue may be prior to obligation.

4.2 Two Notions of Obligation

There are two distinct and important senses of obligation. One of these takes into account the agent's epistemic state, while the other does not.

Suppose, for instance, that unbeknownst to you, there is a hidden button nearby such that if you were to press it, climate change will be drastically mitigated. Ought you to search

for and press this button?

In one sense, yes. Searching for this button is the best thing you can do with your time – by pressing the button you will save millions of future lives and improve countless others.

In another sense, you ought not to search for the button. You would not be *making a mistake* if you did not go searching for the button. For, you lack epistemic access to the fact that there is such a button nearby. Given this lack of epistemic access, it would be irrational for you to search for the button. In addition, you would not be blameworthy if you didn't search for it. On the contrary, you would be blameworthy if you failed to search for the button, if you *knew* it was nearby and had the aforementioned properties.

The first sense of ought is known as the "objective" ought, while the second sense is known as the "subjective" ought.

The subjective ought is the one that depends on the epistemic state of the agent in question. Agents who do not do as they subjectively ought to do are *making a mistake*. This paper attempts to sketch a picture of the kind of epistemic access an agent must have in order to be subjectively obligated to do something.

I claim that the subjective ought bears key connections to other important concepts in ethics. For one, virtuous agents typically do as they subjectively ought to do – they may rarely if ever do as they *objectively* ought to do (think of a virtuous doctor in the 6th century BC). In trying to make progress towards a tenable picture, I will often make use of the second important connection:the subjective ought functions as a *necessary condition* on blameworthiness. You can't be blamed for doing something that you weren't obligated not to do. You can't be blamed for doing something that in no way constituted a mistake.[1] Indeed this seems like a datum. Suppose I say that some agent is blameworthy for an action, and you ask, "What was her mistake?" I say, "She made no mistake in doing what she did." This would not make much sense as a response.

Of course, the subjective ought is not a sufficient condition for blameworthiness. A child

[1]See Harman (2015), Smith (2010), Milo (1984) for further defense of this claim.

may do something she ought not to do in this sense, but nevertheless not be blameworthy – for she has the excuse of being a child.

Keeping in mind that the subjective ought is a necessary condition for blameworthiness will provide additional resources with which to determine subjective obligation in particular cases, and thus will help to arrive at a viable theory.

4.3 Why Moral Beliefs Do Not Matter

Consider the following case.

> ABORTION: Kathy is a teenager who is deciding whether or not to get an abortion. She's fairly confident that if she carries the pregnancy to term, she'll have to give up on several things that are important to her. For instance, given her particular situation, she'll have to give up on going to college and becoming a doctor, which has long been a dream of hers. Moreover, she knows all the relevant medical and biological facts about abortion – the nature of the procedure, the cognitive sophistication of the fetus at various stages, etc. However she is uncertain about whether abortion is permissible.

What should Kathy do? According to *uncertaintists* (e.g. Lockhart (2000) and Sepielli (2009)), what Kathy should do is partly determined by what her moral beliefs (or credences) are. In the case above, then, it could well be the case that Kathy should not have an abortion. However, if she was certain that abortion is permissible, then she should have an abortion.

According to *externalists* (e.g. Weatherson (2013) and Harman (2015)), what Kathy should do depends only on what the *moral facts* are. Thus if abortion is in fact permissible, then Kathy should get the abortion. On the contrary, if it is morally wrong to get an abortion, then Kathy should not get the abortion. Crucially, Kathy's moral beliefs about

whether or not abortion is permissible are *irrelevant* to the question of what she ought to do.

Among other things, uncertaintists are motivated by the idea that norms are meant to be action-guiding. In order to formulate a helpful norm then, we have to take into consideration the agent's epistemic situation with regards to the moral facts – it's not helpful to simply say to the agent: do what is in fact right!

There are several arguments motivating the externalist thesis as well. For one, some authors have claimed that caring about morality *de dicto* is fetishistic. In other words, it's not virtuous to act out of a concern for the instantiation of a moral property for its own sake. What the virtuous agent cares about, non-derivatively, are those things that *ground* moral truths – facts about suffering, pleasure, promise-keeping, etc.[2]

However, this line of argument, I take it, is not decisive. Several authors have argued that caring about morality *de dicto* is not fetishistic.[3] Hence, in this paper I want to focus on an argument that I think is more dialectically powerful. The argument stems from the problem of *misguided conscience*. The worry is that when we consider agents whose moral beliefs are sufficiently mistaken, it will not be plausible that those agents should do what they believe is right.[4] Consider the case below.

> DICTATOR: Mengsk is a dictator who believes that he ought to crush any political opposition that may arise. He believes this because he thinks that democracy would be a morally inferior system, when compared to his rule. Second, he believes that owing to his ethical and economic wisdom, he is morally superior to his subjects – and thus ought to be in a position of power relative to them.

Nonetheless, Mengsk ought not to crush his political opposition, even if he believes this is the right thing to do! If this is correct, however, then it doesn't seem like an agent's *moral*

[2]Smith (1994) makes the point that desiring to do the right thing, where this is read *de dicto*, is fetishistic. See Arpaly (2003) gives a detailed defense of such a view.

[3]See for example Copp (1997) and Olson (2002).

[4]See Harman (2015) for a version of this style of argument.

beliefs are in general what determine what she ought to do. Rather, if we feel a pull in the direction of thinking that, say, what Kathy ought to do in ABORTION can be different from what an agent who is certain that abortion is morally permissible ought to do, then we will have to find an alternative explanation – the explanation can't consist simply in the fact that Kathy is morally uncertain. In this paper, I will propose such an explanation.

Externalists have not aimed to give such an explanation. Rather, they take the problem of misguided conscience to show that in cases like ABORTION, what the agent ought to do depends on what the moral facts are. Hence, Kathy ought to get an abortion if it is in fact morally permissible, and *mutatis mutandis* if it is not.

4.4 Why Non-Moral Beliefs Do Not Matter Either

Externalists thus claim that an agent's moral beliefs are irrelevant for determining what she ought to do. Non-moral beliefs are relevant, however. Therefore while moral uncertainty does not affect what an agent ought to do, factual uncertainty does. Here is a case which might be thought to support the latter claim.

> SUGAR OR ARSENIC: Alice has asked Bob for some sugar in her tea. There are two containers in the cupboard, both filled with what looks like white sugar. Bob knows that one of them is arsenic, but doesn't know which one. What should Bob do?

It seem clear in this case that Bob should avoid using the contents of either container. For, the upside is simply that Alice will enjoy her tea more. But the downside is tremendous – if Bob uses what is in fact arsenic, Alice will get seriously sick and may die.

This might seem to show that an agent's non-moral beliefs play an important role in fixing what she ought to do. At the very least, what we should say about Bob seems more clear-cut than what we should say about Kathy in ABORTION. So perhaps there is a deep

disanalogy between the ways in which moral and non-moral beliefs ought to guide an agent's practical deliberation.

Thus, according to Elizabeth Harman: "A person's moral beliefs and moral credences are usually irrelevant to how she (subjectively) should act. How a person (subjectively) should act usually depends solely on her non-moral beliefs and credences; her moral beliefs and credences are relevant only insofar as they provide warrant for beliefs and credences about what her non-moral situation may be."[5]

Similarly, according to Holly Smith, a principle P that satisfies the following conditional: "*if* agent S believes (correctly or incorrectly) of some act A that A is possible for him to perform and that A has feature F, *then* principle P prescribes A, relative to principle Q and relative to S's non-normative beliefs about A," is a principle of "subjective rightness."[6] Thus, what matters in determining subjective obligation, according to Smith, are the agent's non-normative beliefs; her normative (and hence moral) beliefs are irrelevant.

I think there might be good methodological reasons to reject such views. For, they invite a demand for a satisfactory *explanation* for why non-moral beliefs matter in determining what an agent ought to do, but moral beliefs do not. All else equal, one might argue, we should prefer either a picture on which either all of an agent's beliefs matter or none of them do – for this would lead to a more simple, less disjunctive theory.

The disjunctive picture can be avoided, I think, if we abandon the assumption that the agent's *beliefs* are what matter in determining what she ought to do. This assumption is quite common. Many writers who discuss the "subjective ought" – even if they do not have a stake in the uncertaintism/externalism debate – characterize this ought in terms of the agent's beliefs or credences. Some authors defend the claim that A ought subjectively to Φ just in case A ought objectively to Φ if A's beliefs relevant to the particular situation were true. (This seems to raise the question of what to say about fragmented belief states,

[5]Harman (2015), p58.
[6]Smith (2010), p93.

but let's waive this problem here.) Others leave it more open how the agent's subjective obligations are determined by her beliefs.[7]

However, there are independent, and I think decisive, reasons to reject this view. This is because we can construct a plethora of examples where an agent has some set of mistaken non-moral beliefs, and yet ought *not* to do the action that would be correct if those beliefs were true. Consider the case below.

> MEDICAL SKEPTIC: Martha is skeptical of modern medicine, and in general prefers to go with herbal treatments. In particular, she believes, *pace* what the doctors with MDs say, that a particular Vaccine will put her son Billy at risk for various ailments, and on the other hand, will not be effective in doing what it is meant to do. In fact, Vaccine is crucial to preventing a terrible Disease. Now, Martha lives in modern Western society, and has ample evidence suggesting that Vaccine prevents Disease and has no serious side effects – this is what the news reports medical journals as saying, it's what the medical community says, what the scientists say, etc. Further, the evidence for the claim that herbal treatments can prevent Disease is very weak. For instance, herbal medicines have been around for hundreds of years but Disease was prevalent until only a few decades ago – Martha knows this but, as it were, doesn't put two and two together. Ultimately, Martha avoids vaccinating Billy. When Billy grows up, he gets Disease, as a result of not being vaccinated in time.

In this case, it seems that Martha does not act as she subjectively ought to. She's acting irrationally. This is so despite the fact that *were* her non-moral beliefs about the effects of Vaccine true, she would have been justified in not vaccinating Billy.

Another way to get to this conclusion is to note that Martha seems blameworthy for not vaccinating Billy. She should have known better! Indeed this seems to be an instance where

[7]Among the writers who characterize subjective obligation in this way are Brandt (1965), Prichard (1932), Ross (1939), and Thomson (1986).[8]

the saying "the road to hell is paved with good intentions" resonates. Yet her failure is not a result of having some mistaken moral belief. For if the non-moral beliefs she has about the vaccine were true, then it would have been morally wrong for her to vaccinate Billy – it's wrong to put your child in serious danger.

Now, in the sense of 'ought' we're interested in, it's a necessary condition for it to be the case that A is blameworthy for Φ-ing that A ought not to have Φ-ed. Therefore, Martha ought to have vaccinated Billy even if *were* her mistaken non-moral beliefs true, she ought *not* to have vaccinated Billy. Therefore, what an agent ought to do does not depend on her non-moral beliefs relevant to the matter at hand.

The natural view that the above discussion suggests, I think, is that what an agent subjectively ought to do depends (in some way to be specified) on the agent's *evidence*.[9] Martha has sufficient evidence pointing to the conclusion that the vaccine is effective. Relative to this evidence, she ought to vaccinate Billy – and she's not off the hook simply because she didn't form the correct beliefs based on that evidence.

Recently, Zimmerman (2014) has defended a picture according to which an agent subjectively ought to Φ just in case Φ-ing would constitute the prospectively best option relative to one's evidence *insofar as one has availed oneself of it*. An option is prospectively best, in most cases, according to Zimmerman, when it maximizes *projected* value – i.e. expected value relative to the probabilities and valuations of outcomes that one's evidence supports. Importantly, for Zimmerman, the probabilities relevant here will be determined by the evidence that the agent has availed himself of. Thus, the probabilities relevant to determining the prospectively best option may not match the credences the agent actually has.

Now, Zimmerman adds the qualification that the agent in question needs to have availed herself to evidence because he is worried about the following sort of case (quoted directly):

> Harry, a physician, has a patient, Renée, who has a moderately serious neurolog-

[9]This sort of view has been defended by Dancy (2002), Scanlon (2008), and Kiesewetter (2011), among others.

Option	Actual value	Maximum projected value	Worst possible value
E	?	No	-30
F	?	?	-110
G	?	?	-160

Table 4.1: Harry's Options

ically based tremor. Harry has three options: E (doing nothing), treatment F, and treatment G. Harry knows the possible outcomes of these options, and the values of these outcomes, which are as follows. Each of E, F, and G might result in a cure (+10), and each might result in no cure (-10). In addition, E might result in loss of use of the right foot (-20); F might result in loss of use of the left hand (-100); and G might result in loss of use of the right hand (-150). It is in fact the case that each option would be fully successful...Harry's evidence *does* yield probabilities for the various possible outcomes of his options, and hence these options *do* have a projected value (one that is not simply the result of applying the principle of indifference). Secondly, however, Harry has not managed to arrive at any estimate of these probabilities. Finally, Harry has a well-informed colleague who, while declining to tell Harry which option *would* maximize projected value, has reliably told him that option E would *not* maximize such value. (Zimmerman (2014), p 68-70)

It's important to note that the case doesn't involve an instance of *forgetting*. It's not that Harry has forgotten the relevant evidence about the treatment and the human body. If that's all that were going on, the solution should be easy – since Harry forgot those things, he doesn't have the relevant evidence anymore, which would suffice to fix the probabilities. So of course he should choose E to minimize the risk.

Zimmerman represents the situation from Harry's point of view as shown in Table 4.1.[10] What should Harry do? It seems that he should choose E, even though, given his total evidence, it's not the prospectively best option. Indeed, it would be a mistake to choose either of the other options given the great risk at hand.

In order to accommodate this verdict, Zimmerman says that the evidence that's important for determining subjective obligation is the agent's evidence insofar as he has availed of it. Since Harry has not availed himself of the relevant evidence on the nature of treatments, he should defer to his colleague.

What is the relevant sense of 'avail' here? What has to be the case for an agent to have availed herself of some piece of evidence? Zimmerman suggests that the agent must have formed a relevant *justified belief* based on the evidence. He writes, "one ought to do that act which constitutes the best bet in light of what one is *justified* in believing about one's options."[11] In the case above, Harry has not formed justified beliefs about the possible outcomes of the various treatment options – even if one could in principle do so in his situation.

It's not obvious however, whether Zimmerman's view can accommodate the plausible verdict in the MEDICAL SKEPTIC case. There, Martha does not believe that the vaccine is effective, and *a fortiori* does not *justifiably* believe that it is effective. This raises a dilemma for Zimmerman. On the one hand, he could say that Martha has not availed herself of the evidence – viz., that the doctors say the vaccine is effective, and the like – in that she has not formed the relevant justified belief on the basis of this evidence – i.e. that the vaccine is effective. But this will mean that Martha is doing as she ought to do in refusing to vaccinate Billy. This is implausible. On the other hand, Zimmerman could say that we should think of the relevant justified beliefs as being: the doctors say the vaccine is effective; the media reports journals as saying so, etc. Relative to *these* justified beliefs, failing to vaccinate Billy is not prospectively best, and therefore Martha is making a mistake. However, it seems then

[10] Zimmerman (2014), p69.
[11] Zimmerman (2014), p75.

that the same move could be applied to Harry in the case above. Relative to the more basic pieces of evidence Harry has about the nature of the treatment and the human body, he too is making a mistake in not arriving at any estimate of the probabilities, and *a fortiori* the *right* probabilities – just as Martha is making a mistake in not coming to believe that the vaccine is effective. Thus if Harry gives his patient treatment E, which does not maximize projected value, he would not be doing as he ought to do.

Even supposing that this particular dilemma can be resolved in favor of Zimmerman, there is a lurking problem. The issue is that sometimes one can make a mistake because one doesn't pay *attention* to the relevant evidence. Consider the case below.

> OBLIVIOUS BOB: Bob repeatedly comments on a large pimple that has emerged on Pip's nose. This makes Pip self-conscious, and he clearly displays signs of this self-consciousness – he turns away, subtly covers his nose, etc. Bob is in a position to notice this – however Bob is oblivious and doesn't form the belief that his behavior is causing psychological distress to Pip. Is Bob behaving as he ought to?

Bob is not behaving as he ought to; indeed, Pip may justifiably blame him for his behavior. Bobs failure consists in not paying sufficient attention to the evidence he has – namely that Pip displays the signs of self-consciousness (to be read *de re*), when his pimple is pointed out. The fact that Bob has not availed himself of this evidence, in Zimmerman's sense, seems to be no excuse – that's precisely where his failure lies! Sometimes we can be blamed as a result of failing to notice things.

4.5 Evidentialism About Obligation

The view that OBLIVIOUS BOB militates towards, I think, is that the agent's *total evidence* is what determines what he ought to do. Martha does what she ought not to do because she

acts in a way that is not warranted given her evidence. Likewise Bob acts wrongly, even if he acts in a way that would be permissible relative to his justified beliefs, because he has plenty of evidence that Pip is distressed. Whether or not Bob forms the appropriate belief, namely that Pip is distressed owing to Bob's remarks, in light of this evidence is irrelevant.

What is evidence? There are varying accounts of what an agent's evidence consists in. I want to remain largely neutral to this debate – the task of figuring out what an agent's evidence consists in belongs to the epistemologist. For my purposes, however, it's important to note two things. First, on any plausible view of evidence, an agent's evidence can come apart from what she believes. It's a truism that we can believe something even though our evidence points in the other direction – this is how we make epistemic mistakes. Second, evidence and *justified* belief can come apart in the following sense: one can have evidence that points towards P, but fail to believe P and *a fortiori* fail to justifiably believe P. This is what is happening in the case of OBLIVIOUS BOB.

Let's consider how the second claim would play out according to the view of evidence proposed in Williamson (2000). According to Williamson, evidence = knowledge. Thus an agent's total evidence consists of all the facts that she knows. In OBLIVIOUS BOB, Bob knows that Pip looks so-and-so, covers his nose upon the mention of a pimple, and so on. This evidence, I claim, justifies the belief that Pip is distressed, even though Bob doesn't end up believing this.

Another key feature of our concept of evidence, which will be important in what follows, is that it bears a tight connection with the notion of epistemic justification. Beliefs are justified to the extent that they are based upon one's evidence. This observation is a key constraint than any proper theory of evidence has to meet.

Now, the view that an agent's total evidence is what determines what she ought to do is also able to make sense of commonly discussed cases involving factual ignorance, like SUGAR OR ARSENIC. In this case, part of Bob's evidence is that one of the containers is filled with arsenic. This evidence strongly militates against him using the substance from either

container – for, the risks are so high. Later on I will argue that the view is also able to accommodate our intuitions regarding cases involving false moral beliefs, like DICTATOR, as well as cases involving moral uncertainty, like ABORTION.

Before doing so, however, let me formulate the view under consideration.

> EVIDENTIALISM ABOUT OBLIGATION: For any agent A in circumstances C and total evidence E, whether or not A subjectively ought to Φ is determined entirely by C and E.

What is included in C? C can include various facts about the situation external to the agent – facts like whether her Φ-ing will make people happy, whether she promised to Φ, etc. – as well as facts about A's psychology – e.g. whether she likes Φ-ing. C does not include, however, A's doxastic attitudes; hence the principle says that A's doxastic attitudes are irrelevant in determining whether or not she ought to Φ.

EVIDENTIALISM ABOUT OBLIGATION, I think, allows us to both i) avoid the problems with existing forms of externalism, and ii) capture and explain certain claims that it might seem only uncertaintists are able to make.

As rehearsed earlier, externalists who hold that an agent's subjective obligations depend on her non-moral beliefs face an objection akin to the problem of misguided conscience. The worry is that there will be agents whose non-moral beliefs are sufficiently misguided – i.e. improperly responsive to their evidence – so as to make it implausible that what they ought to do is what is best (in some sense) relative to those beliefs.

But what about somebody who thinks that what an agent ought to do depends only on her non-moral evidence and not her moral evidence (i.e. evidence consisting of propositions with moral content)? This position does not face the problems mentioned in the paragraph above. For, the defender of this view can say all the things we said earlier about MEDICAL SKEPTIC and OBLIVIOUS BOB.

However, this view, in an obvious way, allows one to say that in ABORTION, what Kathy

ought to do depends on whether abortion is permissible. For, the only sort of evidence that is relevant to determining what she ought to do is the non-moral evidence – i.e. the medical and biological evidence. On the other hand, the relevant moral first principles that she may possess as evidence are *irrelevant* in determining whether or not she ought to get an abortion. Importantly, such evidence is irrelevant *even if* it supports believing that abortion is wrong, when it is in fact permissible (and *mutatis mutandis*).

One might think that the externalist should adopt this view, insofar as she thinks that the evidence is what determines obligation.[12] However, this creates a major problem. The problem is that there can, for all we've said thus far, exist cases where an agent's total evidence supports the claim that she ought to Φ and yet she (subjectively) ought not to Φ. Since the ought we are concerned with bears a constitutive connection with practical rationality, it will then be true in such a case that the agent can be *practically rational* in not intending to Φ. However, she will also be *theoretically rational* in believing that she ought to Φ.

Thus, an agent with such a set of total evidence, can be fully rational in at once believing that she ought to Φ and not intending to Φ. In other words, according to the view at hand there can be cases of akrasia which entail *no* irrationality whatsoever on behalf of the agent. Yet, this seems implausible – indeed, throughout the history of ethics, it's been assumed that there is *something* irrational about an akratic state of mind.[13]

Here, I'm assuming that an agent is theoretically rational with respect to a particular belief just in case the belief is supported by her total evidence and the agent is sensitive to these support relations in some sense.[14] Secondly, I'm assuming that if an agent subjectively ought to Φ, then she can be practically rational in Φ-ing. (She will be practically rational in Φ-ing if she does so for the reasons she has to Φ, or alternatively, she Φs in accordance

[12]Harman (2015), for instance, suggests that an agent's moral evidence is usually irrelevant for determining what she ought to do.

[13]Beginning with Aristotle (1999), akratic behavior has been thought to constitute a paradigmatic case of irrationality.

[14]For more on the connection between evidence and theoretical rationality, see Kelly (2014).

with the requirements of rationality.) In other words, there is a certain kind of connection between practical rationality and the subjective ought. Both these assumptions, I take it, are fairly weak and plausible.

In order to avoid the problem sketched above while at the same time maintaining EVIDENTIALISM ABOUT OBLIGATION, one must hold the following view (I defended this in greater detail earlier, in Chapter 2). An agent's total evidence determines what one ought to do in a way such that it also supports one's believing that one ought to do that thing. In other words, the following is true:

> EVIDENCE PRACTICALITY: For any agent A, with total evidence E, E supports A's believing that A ought to Φ if and only if A ought to Φ, given E.

Both 'oughts' are to be read in the subjective sense; the principle is patently false if the ought in question is the objective version. EVIDENCE PRACTICALITY guarantees that any akratic state of mind is irrational. For suppose that A is in an akratic state of mind – he believes that he ought to Φ but does not intend to Φ. Then it must be the case that either: i) his belief that he ought not to Φ is not supported by his evidence, in which case it is theoretically irrational or ii) he intends to Φ when he (subjectively) ought not to Φ, in which case he is practically irrational. Thus an akratic state of mind is guaranteed to exhibit either theoretical irrationality or practical irrationality (or both).

Now, I have said that an agent's beliefs do not determine what she subjectively ought to do. What I mean is that when an agent ought to do something in this sense, it is not (even partly) in virtue of her having some set of beliefs. Rather the fact that she is obliged to do something holds in virtue of her circumstances and her evidence.

However, there can be cases where beliefs are indirectly relevant for determining subjective obligation. This is because sometimes an agent's beliefs can count as part of her evidence. For example, an agent might find herself believing P but not remember the evidential basis on which she came to believe that P. For most, if not all, agents this is evidence

that P is true. Of course, such evidence is defeasible. If the agent were to learn that the belief that P was planted in her mind by a nefarious hypnotist, then she loses the belief-based evidence for P.

4.6 Reframing the Debate Between Uncertaintists and Externalists

If what I have been arguing is right, then I think we should reframe the debate between "uncertaintists" and "externalists." Hitherto, the debate has been framed in the literature as involving a disagreement about whether moral beliefs or moral evidence are relevant in determining what an agent ought to do. But if that's the real debate, then accepting EVIDENCE PRACTICALITY would seem to hand the victory to the uncertaintist. For, according to that thesis, it's the agent's *total evidence* that is relevant in determining what she ought to do – and for all we've said, this may involve moral evidence.

Nonetheless, this is too quick – for we can ask all the important questions, and capture the core distinctions motivating the debate, even supposing that an agent's obligations are determined by her total evidence. Let's go back to the very first case, ABORTION. There, the externalist wanted to say that what Kathy ought to do depends on whether or not abortion is in fact permissible. On the other hand, the uncertaintist wanted to leave it open – even supposing that Kathy knows all the relevant medical and biological facts about abortion – what Kathy ought to do.

Both these possibilities can be accommodated, even supposing the truth of EVIDENCE PRACTICALITY. The externalist verdict is reached if it turns out that Kathy's total evidence supports the belief that she ought to get the abortion, supposing that abortion is in fact permissible. According to EVIDENCE PRACTICALITY it will also turn out that she ought to receive an abortion (given her aims in life).

On the other hand, the uncertaintist verdict is reached to the extent that the available evidence *merits* Kathy's being uncertain as to what she ought to do – for example, if it merits her having a 0.5 credence on both the proposition that she ought to get an abortion and its negation.

The debate between "externalists" and "uncertaintists," then, is one we should reframe as a disagreement about what the evidence supports in various cases. This disagreement can be a result of differing underlying views on what counts as evidence for moral claims, and how that evidence functions in a rational agent.

For example, here is a view that I think delivers externalist verdicts in cases like ABORTION:

> NATURALIZED MORAL EPISTEMOLOGY (NEM): The evidence for moral claims ultimately bottoms out in non-moral facts.[15]

The thought here is that given any moral truth, M, the basic evidence for it will consist of only non-moral truths. Thus for example, consider the claim that punching Billy is wrong. According to NEM, the basic evidence for this claim will only involve facts that do not contain moral properties like badness, wrongness, etc. Hence the evidence for why punching Billy is wrong may well be something like: punching Billy would cause him harm, and causing harm is wrong. However, *ultimately*, the evidence will bottom out in facts about the nature of pain, emotional distress, etc. Thus any person who knows these latter facts will be in an evidential situation such that it's both rational for her to believe that punching Billy is wrong *and* obligatory for her not to punch Billy.

Now, in ABORTION, it's stipulated that Kathy knows the relevant medical and biological facts about abortion. If that's true, then she is in an evidential situation such that she ought to receive the abortion if it's permissible and vice versa. This is because she already has all the evidence relevant to the permissibility of abortion – given this evidence, then, she ought to pursue the abortion if it is permissible.

[15]Setiya (2012), has outlined and defended such a view.

The fact that she's *uncertain* as to whether abortion is permissible is besides the point, given what was said earlier. What matters now is whether she has evidence going in the other direction. What could such evidence consist of? Well, perhaps she has received conflicting testimony. Some wise people say that abortion is permissible, and other wise people say that it is not.

However, we might supplement NEM with a plausible assumption: if the basic evidence for, say, the permissibility of abortion is already in Kathy's possession, then such evidence *defeats* the testimony-based evidence. This is a natural assumption to make because testimony, within the NEM framework, will be providing *indirect* evidence for moral claims, and it's plausible that basic evidence defeats indirect evidence in this scenario.

The case is analogous to the following. Suppose you know that Bill generally leaves campus on the 5 o'clock bus. You see the 5 o'clock bus leaving its station and infer that Bill has gone home. However, when you walk past the department lounge, you spot Bill there and exchange greetings. Your seeing Bill defeats the old, indirect, evidence you had for the claim that Bill went home. The old evidence ceases to have any weight (unless you have some evidence that an excellent impersonator is at large). Likewise, one might think, Kathy's testimony based evidence is defeated by her knowledge of the medical and biological facts.

Whereas, if Kathy had been sufficiently ignorant of the relevant non-moral facts about abortion, then the testimonial evidence to the effect that abortion is morally wrong would have had weight. Indeed, if such evidence were strong enough (if, say, enough wise people said so) and if Kathy were sufficiently ignorant about the relevant non-moral facts, it might have turned out that Kathy ought not to get the abortion.

How might the uncertaintist verdict be reached in the ABORTION case? One way to motivate the verdict is to think of the evidence for the morality of particular acts as consisting partly of the relevant ethical theory, which in turn is supported by first-principles and arguments. There are two main consequences of doing so, with regards to uncertaintism.

First, it's plausible that the usual agent's evidence relevant to Kantianism and consequentialism, for example, merits the agent being uncertain as to which theory is true. In light of this, it might be the case that Kathy is rationally responding to her evidence in assigning 0.5 credence to the permissibility of abortion in the initial case. Ethics is very hard, and extremely smart people who spend their lives studying the field have yet to converge upon one ethical theory.

Second, epistemic permissivism is quite plausibly true if evidence for moral claims works in the way described above. Schoenfield (2014) defines permissivism as "the claim that, sometimes, there is more than one way to rationally respond to a given body of evidence." (p1) In the case of Kathy then, it might be true that given the available evidence, there are multiple rational responses to it. Thus it might be rational for Kathy to assign non-zero credence to the proposition that abortion is impermissible, even if, as a matter of fact, abortion is not permissible. This yields an uncertaintist conclusion as explained earlier.[16]

4.7 From Evidence to Action: Ideal vs Non-Ideal

Here is a possible major worry for EVIDENTIALISM ABOUT OBLIGATION. Consider the following case (represented in Table 4.2):

> EXPERT TESTIMONY: Janet is a high-ranking official in charge of deciding between three economic policies, A, B, and C. Available to her is a large spreadsheet of data that she has looked over, run some models on, etc. Now, Janet is not an expert on the issue – Arvind, however, is. Arvind tells Janet that, according to his analysis, the best policy is B, followed by C, followed by A. Nonetheless, Janet's own modeling says that A is the best policy, followed by B and C. As it turns out, on a fluke, this is also what the evidence supports. But the evidence

[16]For more on epistemic permissivism, see also White (2005) and Kelly (2013).

is tainted – the dataset is deficient in many ways. So, *as a matter of fact*, C is the best policy followed by B, followed by A. What ought Janet to do?

Policy Ranking	Janet	Arvind	Evidence	Actuality
1st	A	B	A	C
2nd	B	C	B	B
3rd	C	A	C	A

Table 4.2: Expert Testimony

Quite plausibly, Janet ought to defer to Arvind here, and implement policy B. Yet, her evidence supports A, in the sense that an ideally rational being with unlimited computational abilities would conclude that A is the best policy, given the spreadsheet data. So there seems to be a problem.

Another way to get at the intuition here is to suppose that Janet implements A, and as a result, many individuals lose their jobs; many of these people would not have lost their jobs had Janet taken up Arvind's suggestion and implemented B. Would Janet be blameworthy for implementing A?

Though opinions may differ on this, I think that it's plausible to say that Janet is blameworthy for implementing A *even if* that is what her evidence supports. In complex cases like these, where the evidence is messy and requires a lot of expertise to analyze satisfactorily, we should defer to the experts. But it seems that according to EVIDENTIALISM ABOUT OBLIGATION, Janet ought to have implemented A, and thus cannot be blameworthy. Does this mean that EVIDENTIALISM ABOUT OBLIGATION is false?

What this example shows, I think, is that we need to distinguish between ideal rationality and reasonableness. According to EVIDENTIALISM ABOUT OBLIGATION, what an agent ought to do is determined by her evidence. But there is still an open question: is what the agent ought to do equivalent to what the *ideally rational* agent (with unlimited computational power and general intelligence) would do given the circumstances? Or should we defend a

less strict version, according to which the agent ought to do what a *reasonable human being* would do given the evidence and circumstances?

Defending the former option commits us to the conclusion that Janet does what she ought (subjectively) to do, and thus is not blameworthy. Accepting the latter option, however, plausibly yields the verdict that Janet *doesn't* do what she ought to do – for a reasonable person in her situation would defer to Arvind, the expert, given the complexity of the evidence. I thus favor this latter option.

4.7.1 Towards a Virtue-Theoretic Conception of the Subjective Ought

What is it to be "reasonable" in the sense given above? Ultimately this is a substantive issue to be decided on various first-order grounds.

However, for two chief reasons among many, I don't think 'reasonable' should just mean 'statistically normal'. First, if there are no constraints on the method of grouping relative to which the agent is statistically normal, it will be drastically indeterminate what an agent ought to do. Second, if statistical normalcy is claimed relative to the agent's culture or milieu, then it will be too easy to get agents off the hook – if your culture's beliefs are bad enough, you can be let off the hook for many actions that are plausibly wrong. Moreover, this view will not be able to make sense of *moral progress*. For suppose an agent does something that exhibits moral courage, by doing the right thing even if the statistically normal agent would not – the view being considered would nonetheless yield that this agent acts in a way that she ought not to. This is extremely implausible.

While what counts as reasonable is an issue for substantive moral debate, I think we can make some initial progress on the matter by taking note of the sorts of capacities that we would deem irrelevant to reasonableness in this sense. For instance, the ability to do complex arithmetic in a matter of seconds does not factor into the extent to which one is reasonable. Thus, for example, suppose that Alice has in front of her all the evidence she needs to figure out how to defuse a ticking bomb. However, she needs to find the answer to

a complex arithmetical sum in five seconds, and she has no calculator on hand. Now, if she fails to find the right answer and therefore fails to defuse the bomb, it doesn't seem as if she did what she ought not to do. For it's not the case that a reasonable person in her situation would have been able to figure it out – humans are not meant to be calculators.

What this kind of example suggests is that reasonableness has to do with the proper exercise of capacities that a well-functioning human would have. In other words, reasonableness has to do with *human virtue*. Indeed, it is not a human virtue (or vice) to be able to perform extremely complex arithmetic in a matter of seconds. This capacity may well be essential to a good *calculator*; but it's not essential to a good *human being*.

These observations point to the following view of subjective obligation:

> VIRTUE THEORETIC SUBJECTIVE OUGHT (VTSO): An agent ought subjectively to Φ in circumstance C given evidence E just in case A's counterpart, A+, who is just like A but lacks the human vices and possesses the human virtues, would Φ in C given E.

It's important that A+ is as close to A as possible while being maximally virtuous. Thus A+ must not be computationally more adept than A, given that arithmetic computational ability (perhaps after a certain point) is not a human virtue.

The view thus avoids the conclusion that Janet in EXPERT TESTIMONY ought to implement A. Janet+, if she doesn't have more computational ability or adeptness in the sort of economic theory that is relevant to the case, would defer to Arvind's testimony. In doing so, Janet+ would display, among other things, the virtue of *intellectual humility*. On the other hand, a view on which Janet+ is given maximal general intelligence and computational capacity says that Janet+ would implement A by reasoning impeccably from the spreadsheet evidence.

This view is importantly distinct from the most common analyses of obligation in terms of

idealized counterparts. Consider, for example, the analysis of ought given in Smith (1994).[17] Smith's is an account of *objective* obligation – his view is that an agent ought to Φ in circumstances C just in case her idealized counterpart would want her to Φ in C. The account I'm interested in presenting and defending here is one of *subjective* obligation. However, there is a natural way to extract an account of subjective obligation from Smith's work. That is, we can add the stipulation that the idealized counterpart has the same evidence as the agent in question. Nonetheless, the resulting picture will run afoul of the problem in EXPERT TESTIMONY, which VTSO is designed to avoid. For, the account would involve counterparts who are fully idealized, and thus have, for instance, unlimited computational power. And Janet's counterpart, if she is ideal in this sense, will want her to choose policy A. Hence, this yields the wrong verdict insofar as we are concerned with subjective obligation.

VTSO also avoids a second major problem, by relativizing reasonableness to individuals. For consider an individual with unusually good computational and mathematical abilities. Such an individual, when faced with an array of complex numerical evidence might be able to correctly determine the course of action that the evidence supports. However, an ordinary reasonable person would not. Nonetheless, it's extremely implausible to say that the computational genius should do what the ordinary reasonable person would do given the evidence! Rather, the genius should act according to her own, correct analysis of the data.

This latter verdict falls out of VTSO. For, the relevant counterpart for the genius is not the ordinary reasonable person, but rather someone who is just like the genius (i.e. has her computational abilities) but is maximally virtuous. Such a person would perform the action best supported by her own analysis of the evidence.

To illustrate, consider again the case of EXPERT TESTIMONY. Suppose Amanda is unusually gifted in mathematics, economics, and data analysis so that she is able to conclude that choosing A is the right course of action relative to the spreadsheet data. In such a case, if Amanda knows this about herself, it seems plausible that she is justified in choosing A rather

[17] Similar accounts are found in Firth (1952) and Railton (1986a), among others.

than deferring to Arvind. This is captured by the fact that according to VTSO, the relevant counterpart for determining Amanda's subjective obligations is not the *ordinary reasonable person*, who would defer to Arvind, but rather Amanda+, who has Amanda's computational and mathematical abilities, and is in addition maximally virtuous. Amanda+ would not defer to Arvind, given her knowledge of her own giftedness – therefore Amanda ought not to defer to Arvind either.

4.7.2 A Problem and Two Solutions for Virtue-Based Analyses of Obligation

There is a well known problem for virtue-based theories of obligation, worked out in detail by Johnson (2003). The problem is that ordinary agents can have obligations to *better themselves*. They also sometimes ought to take steps to mitigate the future effects of their vices, and to seek guidance to correct their moral blind-spots. However, it seems that virtue-based theories of obligation of the sort developed in, for example, Hursthouse (2000) cannot accommodate such claims.

Consider the case of a person with a particular ingrained vice – say, habitual lying. Suppose that the best way to combat this is to seek therapy (which happens not to be too costly, etc.). We should be able to say that this person subjectively ought to seek therapy. But now consider the maximally virtuous counterpart – since he is fully virtuous, by stipulation he will not have the vice of habitual lying. So, he would not seek out therapy for correcting this vice. But if that's right, then what an ordinary agent subjectively ought to do cannot be equivalent to what his maximally virtuous counterpart would do given the circumstances and evidence.

I think there are two ways to respond to this objection, given the purposes of my argument. The first is to employ a strategy used in Railton (1986b) – the idea there is that what's relevant in determining what an agent ought to do is *not* what the idealized counterpart *herself* would do, but rather what the idealized counterpart would want the agent in question to do given her circumstances. Put in the terms of this chapter, the idea would

be that A ought subjectively to Φ just in case A's maximally virtuous (but not otherwise idealized) counterpart with the same evidence would want A to Φ given A's circumstances.

This view is able to accommodate the sort of case described above, where the agent has a particular ingrained vice and ought to seek therapy. For, the maximally virtuous counterpart, given the same evidence as the agent in question, would presumably want the agent to seek therapy, even if the counterpart herself would not seek therapy because he is maximally virtuous and doesn't need it.

The second option is to build the vice(s) into the set of circumstances. In the case of the habitual liar, then, the circumstances he finds himself in include the fact that he is a habitual liar. The maximally virtuous counterpart, then, is to be thought of as a "moment of clarity." The counterpart knows that he is about to revert back into his habitual lying ways once the moment of clarity subsides – what ought he to do? Plausibly, he ought to seek the therapy.

Let me illustrate with another example. Suppose that Klaus's evidence supports his giving 10% of his income to charity. However Klaus has certain indulgent dispositions – once he receives a paycheck, he is disposed to spend it on luxury items rather than giving to charity. But he knows that if he incorporates certain mechanisms of social pressure, he will go through with the donation when the time comes. These strategies might include pledging to give through a certain social group, announcing that he plans to give on Facebook, promising to his close friend that he'll give, etc. Let's suppose that giving to charity is much more morally weighty than the costs associated with the social pressures. Now, we should to be able to say that Klaus subjectively ought to take these steps to ensure that he gives to charity when the time comes. However, we know that Klaus+, the maximally virtuous counterpart, will not have the indulgent disposition in the first place – so how can Klaus's obligations be determined by what Klaus+ would in fact do?

According to the current proposal, however, Klaus+ is just Klaus in a moment of clarity – a moment where all the virtues are present and the vices are gone. Moreover, Klaus+ knows

that the indulgent disposition will be present in the future (this is part of the circumstance he finds himself in). What will Klaus+ do in such a situation? Presumably he will take the steps to incorporate social pressure, as described above. If this is right, then VTSO need not be susceptible to Johnson-style counterexamples, so long as the proviso just described is incorporated.

4.7.3 Reasons-Based Accounts of the Subjective Ought

Some authors have recently suggested that we analyze the subjective ought in terms of *reasons*. However, insofar as they fail to incorporate the notion of virtue, as I have done above, they fail to accommodate cases like EXPERT TESTIMONY.

Consider, for example, the view presented in Schroeder (2009b). Schroeder argues that what you subjectively ought to do is is what you have most *subjective reason* to do.[18] What is a subjective reason? According to Schroeder, "X has a subjective reason to do A just in case she has some beliefs which have the property, if they are true, of making it the case that X has an objective reason to do A." (p 233).

The distinction between subjective and objective reasons is meant to be similar to the one between subjective and objective oughts. If the glass in front of you looks like it contains a gin and tonic, but in fact is filled with petrol, you have subjective reason to take a sip, but no objective reason to do so.[19] However, it can be the case that you have subjective/objective reason to Φ even if you ought not subjectively/objectively to Φ. Thus, suppose you are the designated driver. Then, you ought not subjectively to take a sip. But there may be still be a subjective reason for you to do so, if you enjoy the buzz.

However, analyzing the subjective ought in terms of reasons, by itself, is of no use in avoiding the problem detailed earlier. Consider Janet's situation in EXPERT TESTIMONY. When Janet observes the spreadsheet, we can suppose she comes to believe the key facts

[18]Parfit (2011) gives a similar view on which subjective obligation is determined by *apparent* reasons.

[19]This case is adapted from the famous discussion of "internal" and "external" reasons in Williams (1982). However, the distinction between subjective and objective reasons is different from the one that Williams is interested in examining there.

about the data. But these facts will presumably yield objective reasons the balance of which will favor choosing policy A! Therefore, applying Schroeder's view to this case yields that Janet ought to choose policy A, despite Arvind's contrary advice.

According to Lord (2013), subjective obligation is determined by the reasons you *possess*. What is it to possess a reason, R? Lord proposes, as a necessary condition, that you are in a *position to know* R. (According to Lord, you must also in addition be disposed to treat R as a reason for the action in question.)

But now there arises a dilemma with respect to EXPERT TESTIMONY. Let's suppose Janet is minimally decent so that she is disposed to treat the fact that A is the best economic policy as a (strong) reason to adopt policy A.

The first horn of the dilemma is the following. Lord can say Janet is in a position to know that A is the best policy economically – if she were perfectly rational and had unlimited computational capacity, she could come to know that A is the best policy. In such a case, Lord has to say that Janet ought to choose A – but this is implausible.

Alternatively Lord can claim that Janet is not in a position to know that A is the best policy, because of her particular psychology and expertise. That is, Janet, just as she is, will not be disposed to come to know that A is the best policy given a reasonable amount of time even if she tried. However, this leads to a problem discussed earlier, in MEDICAL SKEPTIC. Martha is not disposed to come to know in this sense that the vaccine is effective – yet, she ought to give Billy the vaccine.

As far as I can see, the only way out is to build virtue in to the notion of 'position to know'. So, the proposal could be something like: A is in a position to know P just in case A+, who is maximally virtuous but computationally and intellectually equivalent to A, would be disposed to come to know P, under the appropriate conditions, given a reasonable amount of time. Such a proposal is resistant to the dilemma above. However, the resulting account of subjective obligation would then not be much different from mine.

4.8 Conclusion

An agent's beliefs, be they moral or non-moral, do not ground her subjective obligations. Rather, what an agent ought to do in the subjective sense depends on what her *evidence* is, along with the particular circumstances she finds herself in.

How does an agent's evidence determine her obligations in this sense? It is tempting to think that an agent ought to do what her ideally rational counterpart would do, holding constant the evidential base and circumstances. Such an account, however, fails to make adequate sense of an important class of cases. On the alternative account I offer, an agent subjectively ought to do something just in case her counterpart who is devoid of human vices and has all the human virtues would perform the said action, given the circumstances and evidential base.

Bibliography

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.

Nomy Arpaly and Timothy Schroeder. *In Praise of Desire*. Oxford University Press, 2013.

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.

Paul Boghossian. What is inference? *Philosophical Studies*, 169(1):1–18, 2014.

Richard Brandt. Towards a credible form of utilitarianism. In Hector-Neri Castaneda and George Nakhnikian, editors, *Morality and the Language of Conduct*. Wayne State University, 1965.

John Broome. Normative requirements. *Ratio*, 12:398–419, 1999.

John Broome. *Rationality Through Reasoning*. Wiley-Blackwell, 2013.

Lewis Carroll. What the tortoise said to achilles. *Mind*, 4(14):278–280, 1895.

David Copp. Belief, reason, and motivation: Michael smith's "the moral problem". *Ethics*, 108(1):33–54, 1997.

Jonathan Dancy. *Practical Reality*. Oxford University Press, 2002.

Jonathan Dancy. *Practical Shape*. forthcoming.

Donald Davidson. Actions, reasons, and causes. *The Journal of Philosophy*, 60(23):685–700, 1963.

Donald Davidson. How is weakness of the will possible? In *Essays on Actions and Events*. Oxford University Press, 1980.

Roderick Firth. Ethical absolutism and the ideal observer. *Philosophy and Phenomenological Research*, 1952.

Harry G. Frankfurt. *The Reasons of Love*. Princeton University Press, 2004.

Gottlob Frege. Logic. In *Posthumous Writings*. University of Chicago Press, 1979.

Joshua Gert. Moral worth, supererogation, and the justifying/requiring distinction. *Philosophical Review*, 121(4):611–618, 2012.

Sven Ove Hansson. Objective or subjective 'ought'? *Utilitas*, 22(1), 2010.

Elizabeth Harman. The irrelevance of moral uncertainty. In Russ Shafer-Landau, editor, *Oxford Studies in Metaethics*. Oxford University Press, 2015.

Gilbert Harman. *Change in View: Principles of Reasoning*. M.I.T. Press/Bradford Books, 1986.

Brian Hedden. Options and subjective ought. *Philosophical Studies*, 2012.

Alison Hills. Moral testimony and moral epistemology. *Ethics*, 120(1):94–127, 2009.

Rosalind Hursthouse. *On Virtue Ethics*. Oxford: Oxford University Press, 2000.

Robert N. Johnson. Virtue and right. *Ethics*, 113(4):810–834, 2003.

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 can be permissive. In John Turri Mathias Steup and Ernest Sosa,

editors, *Contemporary Debates in Epistemology (2nd Ed)*. Blackwell, 2013.

Thomas Kelly. Evidence. *Stanford Encyclopedia of Philosophy*, 2014.

Benjamin Kiesewetter. 'ought' and the perspective of the agent. *Journal of Ethics and Social Philosophy*, 2011.

Sebastian Köhler. Do expressivists have an attitude problem? *Ethics*, 123(3):479–507, 2013.

Saul Kripke. *Wittgenstein on Rules and Private Language*. Harvard University Press, 1982.

David Lewis. Finkish dispositions. *The Philosophical Quarterly*, 47:143–158, 1997.

Ted Lockhart. *Moral Uncertainty and Its Consequences*. Oxford University Press, 2000.

Errol Lord. *The Importance of Being Rational*. PhD thesis, Princeton University, 2013.

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.

Ron Milo. *Immorality*. Princeton University Press, 1984.

Jonas Olson. Are desires de dicto fetishistic? *Inquiry*, 45(1):89–96, 2002.

Derek Parfit. *On What Matters*. Oxford University Press, 2011.

Philip Pettit. Joining the dots. In Frank Jackson Geoffrey Brennan, Robert Goodin and Michael Smith, editors, *Common Minds: Themes from the Philosophy of Philip Pettit*. Oxford University Press, 2007.

Plato. *Complete Works*. Hackett Publishing Company, Inc., 1997.

H.A. Prichard. *Duty and Ignorance of Fact*. London: Humphrey Milford, 1932.

Peter Railton. Facts and values. *Philosophical Topics*, 1986a.

Peter Railton. Moral realism. *Philosophical Review*, 95(2):163–207, 1986b.

Gideon Rosen. Metaphysical dependence: Grounding and reduction. In Bob Hale and Aviv Hoffmann, editors, *Modality: Metaphysics, Logic, and Epistemology*. Oxford University Press, 2010.

W.D. Ross. *The Foundations of Ethics*. Oxford University Press, 1939.

T.M. Scanlon. *Moral Dimensions: Permissibility, Meaning, Blame*. Belknap Press of Harvard University Press, 2008.

Samuel Scheffler. *Equality and Tradition*. Oxford University Press, 2010.

Markus Schlosser. Basic deviance reconsidered. *Analysis*, 67(3):186–194, 2007.

Miriam Schoenfield. Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, 48(2):193–218, 2014.

Mark Schroeder. *Slaves of the Passions*. Oxford University Press, 2007.

Mark Schroeder. Buck-passers' negative thesis. *Philosophical Explorations*, 12(3), 2009a.

Mark Schroeder. Means-end coherence, stringency, and subjective reasons. *Philosophical Studies*, 143(2):223–48, 2009b.

Eric Schwitzgebel. Belief. *Stanford Encyclopedia of Philosophy*, 2010.

Andrew Sepielli. What to do when you don't know what to do. *Oxford Studies in Metaethics*, IV, 2009.

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

Kieran Setiya. What is a reason to act? *Philosophical Studies*, 167(2):221–235, 2014.

Holly Smith. Subjective rightness. *Social Philosophy and Policy*, 27(2):64–110, 2010.

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. Imposing risks. In William Parent, editor, *Rights, Restitution, and Risk*. Harvard University Press, 1986.

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

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

Ralph Wedgwood. The normative force of reasoning. *Noûs*, 40(4):660–686, 2006.

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

Roger White. Epistemic permissiveness. *Philosophical Perspectives*, 2005.

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

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

Ludwig Wittgenstein. *Philosophical Investigations*. Oxford: Basil Blackwell, 1953.

Crispin Wright. Comment on paul boghossian, "what is inference?". *Philosophical Studies*, 169(1):27–37, 2014.

Michael Zimmerman. *Ignorance and Moral Obligation*. Oxford University Press, 2014.