

ACCEPTING IT

by

Ian Evans

---

A Dissertation Submitted to the Faculty of the

Department of Philosophy

In Partial Fulfillment of the Requirements  
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2012

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Ian Evans entitled Accepting and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

\_\_\_\_\_ Date: 11/27/2012  
Stewart Cohen

\_\_\_\_\_ Date: 11/27/2012  
Juan Comesaña

\_\_\_\_\_ Date: 11/27/2012  
Keith Lehrer

\_\_\_\_\_ Date: 11/27/2012  
Michael McKenna

\_\_\_\_\_ Date:

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

\_\_\_\_\_ Date: 11/27/2012  
Dissertation Director: Stewart Cohen

### STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Ian Evans

### STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Vijay Chamin Agarwal

## ACKNOWLEDGEMENTS

I owe a debt of gratitude to Nathan Ballantyne, for encouraging me to pursue this strange and unwieldy project, and, more importantly, for his friendship throughout the difficult years of graduate school. Stewart Cohen taught me never to settle for an obscure thought or well-used metaphor – absolute clarity or nothing at all. Juan Comesaña’s knack for devastating objections kept me awake at night, and improved my argumentation substantially. I would like to thank Michael McKenna for his generosity, which is matched by his willingness to tell me when the work did not meet his expectations and to encourage me to improve. And, of course, there’s Keith Lehrer, who was kind enough to let me stand on his shoulders.

Thanks are also due to my dream team of undergraduate advisors: Nicholas Smith, Rebecca Copenhaver, Jay Odenbaugh, J.M. Fritzman, and William Rottschaefter. I would never have made it to graduate school without their unflagging support.

My parents Chris and Debbie Evans, and my sister Dana Evans, have always been my anchors. I am grateful for their love and support. Finally, I would like to thank Murphy Dwyer, my shelter in the storm.

## DEDICATION

To Lamont Coleman; rest in peace.

## TABLE OF CONTENTS

ABSTRACT.....	7
CHAPTER 1: REVEALED IRRATIONALITY.....	8
CHAPTER 2: THE DISPOSITIONALIST GAMBIT .....	46
CHAPTER 3: ACCEPTANCE, ALIEF, AND IN-BETWEEN BELIEF .....	79
REFERENCES.....	114

## ABSTRACT

Most of us are aware of believing things we know we should not – often, we cannot help it. My dissertation comprises three papers exploring this phenomenon. In “Revealed Irrationality,” I defend the possibility of believing, of one of one’s beliefs, that it is irrational. “The Dispositionalist Gambit” examines a recent strategy for defending the claim that we have voluntary control over what we believe and finds it wanting. But there is a propositional attitude we have control over – what I call *acceptance* – and in “Acceptance, Alief, and In-Between Belief” I argue that we need this attitude to explain otherwise puzzling behavior.



## CHAPTER 1: REVEALED IRRATIONALITY

Ordinarily, coming to believe that something you believed is irrational leads you to abandon the belief. But could you come to believe that one of your beliefs is irrational, yet persist in believing it anyway? Could you coherently think that something you now believe – and will continue believing for the foreseeable future – is irrational or unsupported by your evidence? Might some beliefs be so “sticky” that higher-order beliefs about their rationality have no force over them? What’s at issue is not whether some beliefs are recalcitrant in the face of countervailing evidence. Most everyone agrees that some beliefs are like that. The question is whether you could recognize such a belief of your own, while going on believing it and without self-deception. Let’s such a conflicted state, “revealed irrationality.”

Many philosophers say “no”: they think it incredible to suppose you could recognize that you irrationally believe *p*, while continuing to believe *p*. These philosophers think that such a situation is ruled out by the very nature of belief. If one *really* believed that it was irrational to believe *p*, one would no longer believe *p*.<sup>1</sup> One hears such claims in the hallways, but the arguments behind them are not often discussed. So I went in search of the arguments. The aim of this paper is to report my findings. In brief: there are several arguments that have been influential, but none withstand scrutiny.

---

<sup>1</sup> Or perhaps a slightly weaker claim is intended: if one really believed that it was irrational to believe *p*, one would no longer believe that one believed that *p*, due to some sort of self-deception.

Of course, merely blocking these arguments is unlikely to alleviate the incredulity with which many philosophers regard revealed irrationality. With that in mind, the second task of this paper is to explain why revealed irrationality is possible. If I'm successful, I'll have showed not only that no extant argument against revealed irrationality succeeds but also why no such argument *can* succeed.

### **Examples**

Let's begin with an example, so that we have something concrete in mind as we think about the arguments to come. Brian Ribiero reports:

There are well-known skeptical arguments which purport to show that [believing that I have two hands] is not justified. These arguments purport to show that, instead of believing that I have two hands, I ought to suspend judgment about whether I have two hands, if I approve of the skeptical argument(s). ... I do judge that there is a completely persuasive argument for radical skepticism. I do reflectively accept and (even publicly) endorse that argument. As far as I can tell, my acceptance of that argument's cogency is as clear-headed as my acceptance of anything is. I spend considerable time thinking about the argument, writing about it, defending it at home and abroad. I think that skeptical argument is a winner. Period. But I still believe I have two hands. And that is the problem. (2011: 21)

Further, he implores us to imagine what would happen if we were in his shoes:

Suppose that you were to come to accept the cogency of some radically skeptical argument. Surely this could happen to you — even many anti-skeptics readily acknowledge the power of skeptical argumentation. Now suppose it did happen: do you suppose you would stop believing that you have two hands and start suspending judgment about the matter? Don't be silly: you'd do no such thing. (22)

So, according to Ribiero, he believes that he has two hands. But he also believes that he shouldn't believe this: that a powerful skeptical argument shows that he should suspend judgment on whether he has hands. But he can't help but continue believing he has hands. What's more, he says that we'd all do the same were we in his shoes. Thomas Reid made essentially the same observation in his famous ribbing of philosophical skeptics:<sup>2</sup>

Pyrrho of Elis, the father of [skepticism], seems to have embraced it more thoroughly than any of his successors; for it is reported ... that his life corresponded to his doctrine. Thus, if a cart ran against him or a dog attacked him ... he wouldn't stir a foot to avoid the danger, giving no credit to his senses. Luckily for him he had servants who weren't such great sceptics; they took care to keep him out of harm's way, so that he lived to be ninety years old.

Though his followers no doubt accepted Pyrrhonian skepticism, they couldn't help but believe, for example, that they had hands or, in the right circumstances, that Pyrrho was being attacked by a dog; they couldn't get their first-order beliefs in accord with their skepticism. Reid continues:

[T]he great Pyrrho himself sometimes forgot his principles. He is said once to have been in such a rage with his cook – probably for not roasting his dinner to his liking – that he chased the cook even into the marketplace, holding the spit with the meat on it.

Reid's point here, of course, is that *of course* Pyrrho himself could not stop believing his senses: such beliefs are *automatic* in us. So the Pyrrhonians, according to Reid, were in

---

<sup>2</sup> Reid's point, of course, was to show the absurdity of radical skepticism. We can leave the debate between Reid and the skeptics for another day.

the unfortunate position of, for a great many  $p$ , thinking they shouldn't believe  $p$  (even actively resisting) but going on believing  $p$  anyway. Reid alleges that Humean skeptics and Berkeleyan idealists were in the same position. And, indeed, Hume himself seems to have admitted this tension between what his philosophy committed him to and what he couldn't help believing when he returned "to the billiard room."<sup>3</sup>

So Ribiero is another in a long tradition of philosophical skeptics who seem to believe that many of their own common sense beliefs are irrational. Now a defense of the claim that these skeptics are in such a doxastic state would require an argument that attributing these states to them is the best overall explanation of their puzzling behavior. That is not the project of the present paper.<sup>4</sup> Here we examine the prior question whether such a doxastic attribution could *possibly* be correct of *any* agent. To pursue this question, we'll need to be clearer about what doxastic state is in question. Now, skeptics typically (claim to) believe some global skeptical thesis:

*Skepticism*: It is irrational to believe anything about the external world.<sup>5</sup>

---

3 For a lively discussion of a related phenomenon, see Elga (2005).

4 Parts of that project are undertaken in my "Acceptance, Alief, and In-Between Belief" (unpublished a), though my concern there is with subjects who have a slightly different behavioral profile.

5 Strictly speaking, Pyrrhonian skepticism is a slightly different beast. Rather than adopt some skeptical thesis, Pyrrhonians claim to employ a special method that always leads them to doubt. Whether Pyrrhonian skeptics end up with the sort of conflicted beliefs that Reid attributes to them is a delicate (but interesting!) issue that I'll have to leave for another day. But no matter – plenty of skeptics fit my mold.

But some of them think, not implausibly, that they nonetheless believe many things about the external world, such as:

- I have two hands.
- Pyrrho is being attacked by a wild dog.
- My cook did not roast my dinner to my liking.

It is unlikely that these skeptics believe, of each of their beliefs about the external world, that it is irrational: they surely haven't formed such a meta-belief for each of their many external world beliefs.<sup>6</sup> Indeed, some skeptics probably believe that they don't have any external world beliefs.<sup>7</sup> But if they're reflective, most will have formed some such meta-beliefs. For they will have occurrently believed something like "I believe I have hands" and noticed that it follows from this and their skeptical thesis that "I irrationally believe I have hands," which conclusion they allege to believe.

Let's imagine, step by step, a skeptic getting into this situation:

(T) The skeptic forms the belief that she believes p.

---

<sup>6</sup> Perhaps it is right to say that they tacitly believe this of each of their first-order beliefs; but no one has ever been able to come up with a plausible account of tacit belief, so we need not pursue the question here.

<sup>7</sup> Apparently ignoring Socrates exhortation to "know thyself!"

(T+1) The skeptic forms the belief that it is irrational to believe *p* (she derives this from her skeptical thesis).

(T+2) So the skeptic forms the belief that she irrationally believed, at *t*, that *p*.<sup>8</sup>

That that much can happen is not controversial – that’s how ordinary conscious belief revision gets going. But what allegedly happens next *is* controversial:

(T+2) The skeptic continues to believe that *p*.

(T+2) The skeptic continues to believe that she believes *p* (and is unlikely to stop anytime soon).

(T+2) The skeptic continues to believe that it is irrational to believe *p*.

(T+3) The skeptic forms the belief, of her own current belief that *p*, that it is irrational and that she continues to believe it.

The arguments we will consider try to show that the skeptic could not harbor these latter three beliefs at (T+2) and so that is why the state described at (T+3) is not possible.

What will these philosophers say about our skeptics – about Ribiero, for example?

They must say that Ribiero is confused. Either he doesn’t really believe it is irrational to

---

<sup>8</sup> We have in mind, of course, a case where the skeptic still believes at *t+1* that it is irrational to believe *p*. This is in contrast to a case where she gets some new evidence at *t+1* and thinks: “I was irrational to believe it, but now I’m not.”

believe he has hands or else he doesn't really believe that he *believes* he has hands. The idea behind this second option is that Ribiero is confused about the nature of belief so that the attitude that he ascribes to himself with the content "I have hands" is not belief – the concept Ribiero associates with the word "belief" is not belief. This, I think, is what my opponents would say about Ribiero and other skeptics like him.

Before turning to arguments, I want to make two final clarificatory points. First, the notion of irrationality in play may require some explication. The rationality at issue is *epistemic* as opposed to *practical*. Epistemic rationality is, on my usage, the rationality of belief. That's analytic. What's not analytic but, I think, true is:

*Evidentialism*: Only evidence for p can make believing p rational.<sup>9</sup>

Belief is a state and the only thing that rationalizes it is evidence. So, while a desire to please my peers might make it rational for me to *try to get myself* to believe that U2 is a good band, that desire does not make it rational for me to *believe* that U2 is a good band.<sup>10</sup> Trying to get oneself to believe something is an *action* governed by practical rationality; believing something is a *state* governed by epistemic rationality.

---

<sup>9</sup> Sometimes, I think, it takes an argument to move from evidence for p to rational belief that p (the evidence serving as some of the premises in the argument). So there may be a sense in which something besides evidence sometimes makes belief rational. Let's leave the case of apriori rational beliefs aside — I'm not defending an account of rationality, just gesturing at what I mean.

<sup>10</sup> Though there are strange cases where a desire for p is evidence for p, I mean this not to be one of them.

So, what's at issue, really, is whether it's possible to believe that one believes  $p$ , while also believing that one lacks sufficient evidence for  $p$  – all at the same time, in full awareness. This conception of epistemic rationality seems to be common ground between my opponents and I: their arguments against the possibility of revealed irrationality all hang on claims about the relationship between belief and evidence. You may disagree with this conception of epistemic rationality, but if you do, then you are on my side in the present dispute: you ought to think it's possible to believe that your belief that  $p$  lacks sufficient evidence because you might have other, non-evidential, reasons to believe  $p$ . The interest of the dispute, then, is not limited to those agree with me about epistemic rationality.

One final point to further motivate the discussion. It's not obvious that the state of revealed irrationality – if possible – need always be irrational. Certainly, the following looks like a plausible candidate for a norm of rationality:

*Meta-Coherence*: One ought not (believe  $p$  and believe that believing that  $p$  is irrational)<sup>11</sup>

On reflection, however, it's not obvious why it should be correct. Suppose one is in the conjunctive state *Meta-Coherence* prohibits: one believes  $p$  and one believes that believing  $p$  is irrational (on one's evidence, or whatever). Surely it's possible to falsely

---

<sup>11</sup> The parentheses indicate that the "ought" is intended to take wide-scope in this norm, on the assumption that it's possible to make sense of a wide-scope reading of epistemic oughts.



believe that believing p is irrational on one's evidence. But if that's possible, might one not *rationally*, but falsely, believe that believing p is irrational? In general, being false is not sufficient for being irrational, so it's not obvious that it should be in this special case.<sup>12</sup> If that's possible, then it's possible to rationally believe p *and* rationally believe that it is irrational to believe p. In such a case, rationality requires believing p and believing that such a belief is irrational. But that is just to say that rationality requires revealed irrationality in this case. If revealed irrationality is sometimes a requirement of rationality, then it had better be possible! So those who want to deny the possibility of revealed irrationality have another burden to shoulder: they must show that one's reasons could never support both that p and that believing p is irrational.

With all that throat-clearing behind us, let us, finally, turn to the arguments against revealed irrationality. We begin with an argument from Petit & Smith.

### **Petit & Smith's Transparency Argument**

---

<sup>12</sup> Titelbaum (unpublished) has argued that mistakes about rationality are mistakes of rationality: if you incorrectly believe that p is irrational, then you irrationally believe that p is irrational. He may be right about this, but it's certainly not obvious.

In an article devoted to defending that we have the ability to respond to reasons in how we believe and desire, Petit and Smith (1996: 448) attempt to argue that we cannot recognize that one of our beliefs is contrary “to the requirement of fact and evidence.”<sup>13</sup>

Imagine that your beliefs run counter to what evidence and fact require. In such a case, your beliefs will not allow those requirements to remain visible because the offending beliefs themselves give you your sense of what is and your sense of what appears to be. You are therefore denied an experience whose content is that you are believing such-and-such in defiance of the requirements of fact and evidence.

Petit & Smith say that a belief “runs counter to what fact requires” if it is false. They don’t explain what “evidence requires,” but let us agree that one’s belief runs counter to what evidence requires if it is not adequately supported by one’s evidence. So, Petit & Smith argue that one cannot believe one of one’s own beliefs to be false and also that one cannot believe one of one’s own beliefs to be inadequately supported by one’s evidence. It is the second thesis that concerns us here: that is a denial of the possibility of revealed irrationality. The argument draws this strong conclusion from a single – and rather obscure – basic premise:

1. Your belief that p “gives you your sense of what is and what appears to be.”

---

<sup>13</sup> They offer this as an explanation of why so many thinkers have recognized our capacity for free will while simultaneously failing to recognize our capacity for free belief. When we fail to exercise our free will, we tend to be keenly aware of going against what practical reason requires of us. But this same awareness, they claim, is not possible when we fail to exercise free thought. Hence, free thought is “essentially elusive.” No wonder so many have denied its existence.

2. Therefore, you cannot believe that your belief that p runs counter to the requirements of fact and evidence.

How shall we interpret the first premise? I take it that “one’s sense of what is” pertains to the requirements of fact, while “one’s sense of what appears to be” pertains to the requirements of evidence. Let’s focus on the claim that your belief that p gives you your sense of “what appears to be.” We can read that as the following conditional: if one believes that p, then “one’s sense” is that the evidence supports p. If we interpret the first premise in this way, then there seems to be a fairly direct route to denying the possibility of revealed irrationality:

1. Suppose, for conditional proof, that S believes p.
2. If S believes p, then S believes that the evidence supports p. [Petit & Smith’s principle]
3. S cannot believe p and believe  $\sim$ p. [Assumption that contradictory beliefs are impossible.]
4. S does not believe that the evidence doesn’t support p. [From 1, 2, and 3]
5. If S believes p, then S doesn’t believe the evidence doesn’t support p. [By conditional proof on 1-4.]

This shows that S cannot believe p and also believe that the evidence doesn't support p. But notice that, by itself, this doesn't entail that S cannot believe "I believe p, but the evidence doesn't support p" (merely having such a belief seems to be a relatively weak form of revealed irrationality). One might mistakenly believe one believes p, in which case S would not both believe p and that the evidence doesn't support p. Or S might believe "I believe p, but the evidence doesn't support p" without believing "the evidence doesn't support p." So to rule out even weak forms of revealed irrationality, the argument needs to assume:

(Luminosity) If S believes that S believes that p, then S believes that p.

And one of either

(Conjunction Elimination) If S believes "p & q" then S believes p and S believes q

or

(Strengthened Contradiction) S cannot believe p and believe ( $\sim p$  & q)

I suppose strengthened contradiction is as plausible as the standard prohibition against contradictory beliefs (i.e., not very!). But Williamson-style arguments show that this sort of unrestricted Luminosity principle is false. So Petit and Smith's argument leaves open the possibility of believing "I believe p, but the evidence doesn't support p," in cases where one does not actually believe p. A kind of revealed irrationality survives even if we

grant Petit and Smith their premises. It is, however, a *weak* kind of revealed irrationality: the phenomena we started with seemed to involve subjects believing *of* some of their beliefs *that* they're irrational. Showing that such a *de re* belief is impossible would be no mean accomplishment, so we should assess the plausibility of Petit and Smith's premises.

We might plausibly quibble about the possibility of contradictory beliefs, but the key premise in the argument is the first one: if one believes that *p*, then one believes that the evidence supports *p*. Plainly, this conditional needs weakening. Not all believers possess a concept of evidence. Further, not every belief is accompanied by a meta-belief about the evidence for that belief, on pain of regress (and on pain of implausibly overintellectualizing believers). Perhaps something like the following will do:

If *S* believes *p*, and *S* has reflected on whether the evidence supports *p*, then *S* believes that the evidence supports *p*.

Following the literature, we might call this principle *Transparency*, and capture it in the following slogan: the question whether one's evidence supports *p* is transparent to the question whether *p*.

There is a fairly obvious problem with trying to use this principle in arguing against the possibility of revealed irrationality: it is plainly question-begging in such a context. We began with what look like *counterexamples* to this principle. The onus is on the advocate of this sort principles to defend it (or at least show that the counterexamples

fail). Petit and Smith, however, say nothing in the way of defending their Transparency thesis. So, while they may be right, and while there is a route from the Transparency thesis to the denial of revealed irrationality, Petit and Smith have not successfully defended their thesis.

I turn now to an argument from Jonathan Adler that purports to establish a version of the Transparency thesis. Adler's argument, if successful, would vindicate Petit and Smith.

### **Adler's Theoretical Reasoning Argument**

Jonathan Adler (2002) has argued that revealed irrationality is impossible due to the nature of theoretical reasoning (and that, in this way, theoretical reasoning is importantly different from practical reasoning). Let's begin with what Adler takes to be his chief insight, an alleged fact about theoretical reasoning (4):

The objective of theoretical reasoning is to determine whether a hypothesis is true, not whether it is best supported.

According to Adler, when engaging in theoretical reasoning about some proposition  $p$ , we take stock of our evidence in order to determine whether  $p$  is true. Noticing that, on our evidence,  $p$  is more probable than not is insufficient to rationally conclude  $p$  – after all, that  $p$  is more probable than not doesn't guarantee that  $p$  is all that probable. So if it aims

at what is true, theoretical reasoning cannot rest content with conclusions about what is best supported on our evidence. Theoretical reasoning requires sufficient evidence to draw an all-out conclusion of either  $p$  or not- $p$ ; otherwise, it must fail to conclude anything. (This is supposed to be importantly different from practical reasoning where the aim is to determine which action is best given my reasons. For our purposes, we can safely ignore issues about practical reasoning and/or akrasia.) We're now in a position to reconstruct Adler's argument against the possibility of epistemic akrasia:

1. To recognize one believes  $p$  in defiance of one's evidence, one must engage in theoretical reasoning about  $p$ .
2. Theoretical reasoning about  $p$  – if it reaches any conclusion at all – must terminate in one's concluding either  $p$  or not- $p$ .
3. If one concludes  $p$ , then one does not recognize oneself to be believing  $p$  in defiance of one's evidence.
4. If one concludes not- $p$ , then one thereby believes not- $p$ , and so no longer believes  $p$  (and hence one does not recognize oneself to believe  $p$  in defiance of one's evidence).
5. One cannot recognize that one believes  $p$  in defiance of one's evidence.

This, at any rate, is my best attempt at organizing Adler's many interesting claims into an argument for the impossibility of akratic belief. There are three problems with this argument: the first premise is false, the fourth premise is false, and the argument is invalid.

The first premise is supposed to be supported by Adler's conception of theoretical reasoning. I wish to grant Adler this conception: it seems to me that premise 1 is still false. If the hypothesis in consideration is  $p$ , then Adler is correct that theoretical reasoning about  $p$  won't be able to terminate merely in a conclusion about whether  $p$  is best supported by one's evidence. Now if theoretical reasoning were limited to first-order hypotheses, then the only way to conclude that  $p$  was not supported by the evidence would be to reason about  $p$  and conclude not- $p$ . I think that is what Adler has in mind. But theoretical reasoning is *demonstrably not* limited to first-order hypotheses: I can investigate the hypothesis that  $p$  is best supported by the evidence. Call that hypothesis  $q$ . If I engage in theoretical reasoning about  $q$ , Adler holds I either conclude  $q$  is true or false, or else fail to reach a conclusion — I can't conclude that  $q$  is best supported by the evidence. But to conclude  $q$  is true (or false) is to conclude that  $p$ , some first-order hypothesis, is best supported by the evidence. So the first premise is simply false. One way to conclude that one believes  $p$  in defiance of the evidence is to reason, not about the hypothesis  $p$ , but about the hypothesis that  $p$  is best supported by the evidence. The only way that Adler can block this is by adopting an implausibly strong account of theoretical



reasoning according to which investigating directly about what the evidence supports is impossible (an account on which epistemology itself would seem to be impossible). It may be true that we don't *typically* engage in this sort of second-order reasoning, but that we *can* is not up for dispute; and it is precisely the sort of reasoning that leads the skeptics with which we began into their conundrum.

I also think the fourth premise of the argument is false. We'll see why more fully later, but for now I'd like to cast some doubt on it. It seems to me that I frequently conclude things in reasoning that I don't end up believing: often, the conclusion seems *unbelievable*, as in the case of a paradox. I've reached the conclusion many times that no one has free will, but I'm pretty sure I've never really believed it. Adler does, of course, attempt to defend this central claim (2002: 4):

One cannot openly judge the conclusion of one's theoretical reasoning to be *p*, and fail to accept – fully believe – *p*. For belief is nothing but an attitude that the content of the belief is true.

The idea, then, is that concluding that *p* is an attitude that *p* is true, and belief is nothing over and above such an attitude.<sup>14</sup> But this is obscure. What is “an attitude that *p* is true”? That is an undefined technical locution that we don't have any pretheoretic grasp on if. If it just means “a belief that *p*” (which I all I can make of it) then we don't have an argument that one cannot conclude *p* but continue to believe not-*p*. Absent some

---

<sup>14</sup> Adler's argument here requires the further premise that contradictory beliefs are impossible. We'll leave that aside.

substantive account of “an attitude that p is true” that vindicates Adler, it seems that we should take appearances at face value: the conclusions of reasoning do not always generate corresponding beliefs.

But even if Adler is right about these two premises, his conclusion does not follow. The argument purports to take the form of a dilemma: one’s theoretical reasoning must either lead one to conclude p or not-p. But this, of course, is a false dilemma and by Adler’s own lights. For if one deems the evidence inadequate for establishing either p or not-p, then one will fail to draw any conclusion. But then what is to prevent one from continuing to believe p? Even if Adler is right that a conclusion from theoretical reasoning just is a belief (and that contradictory beliefs are impossible), this says nothing about what happens when theoretical reasoning fails to reach a conclusion. In that case, there is nothing in Adler’s machinery to prevent the subject from continuing in her belief that p. So, for all Adler has said, here is a possible way one could come to recognize that one’s evidence is inadequate for believing p, but nevertheless continue to believe p.

I conclude that Adler’s argument fails: the possibility of revealed irrationality still stands. But before leaving Adler, it will be instructive to look at the way he summarizes his argument (2002: 18):

To review: The fundamental disanalogy is that the goal of [theoretical reasoning] is all-out or full belief, and so the (threshold) acceptance of a proposition. When [theoretical reasoning] reaches that goal, contrary or undermining evidence is

nullified. So there is no evidence to play the role of conflicting desires in drawing the agent away from his better judgment.

This reasoning — about nullifying evidence — points to a different way of arguing against revealed irrationality, one that Susan Hurley has developed and to which we now turn.<sup>15</sup>

### **Hurley's Subsumption Argument**

One of the projects Susan Hurley tackles in her (1989) is to explain what she sees as an important difference between practical and theoretical reasons. The difference concerns the notion of *akrasia*: willfully  $\phi$ -ing against one's own better judgment. Practical *akrasia* is deliberately acting in a way that one judges one ought not, all things considered, act. Epistemic *akrasia* is "deliberately" *believing* something that one judges one ought not, all things considered, believe. It is more or less uncontroversial that akratic action is possible and much ink has been spilled trying to make room for it in our theories of action. Hurley is one of the first to have explicitly asked whether *epistemic* *akrasia* is possible and, like most who have since written on the topic, her answer is "no."

Before we examine her argument, a few remarks about this notion of epistemic *akrasia* are in order. While this is not always made clear in the literature, there are two conditions on akratic belief – what Owens (2002) has usefully called *the control condition*

---

<sup>15</sup> Though Adler's characterization here has the same defect as the argument of his we just considered: it fails to consider what happens when theoretical reasoning *fails* to reach a conclusion.

and *the judgment condition*. An akratic belief is, of course, one that the believer judges to be epistemically irrational – this is the judgment condition. But an akratic belief is also one that is, in some sense, in the subject’s *control*: akratic belief is free belief. What I am calling revealed irrationality, then, is weaker than akratic belief, for it has no control condition. But Hurley’s arguments attack the possibility of a belief satisfying the *judgment condition*, so they attack the possibility of revealed irrationality. Some of the other authors we will discuss have also set their sights on epistemic akrasia; like Hurley, though, their arguments attack the judgment condition.<sup>16</sup> From here on out, then, I will write as though they are all primarily concerned with my notion of revealed irrationality.

Now consider the following situation of conflicting practical reasons:

Someone believes that  $r_1$  is a reason to  $A_1$ , while  $r_2$  is a reason to  $A_2$ . She believes, further, that she has an all-things-considered reason to  $A_2$  (weighing  $r_1$  against  $r_2$ ). Could she, nonetheless, go on to  $A_1$  on the basis of  $r_1$ ?

The answer, of course, is “yes.” Though  $r_1$  is outweighed by  $r_2$ , it is still a reason to  $A_1$  and so she could intelligibly, if irrationally, be motivated by it. But consider a contrasting case of conflicting *epistemic* reasons:

---

<sup>16</sup> This is typical of the literature on epistemic akrasia. The one exception I am aware of is Owens himself: he agrees with me that beliefs can satisfy the judgment condition, but argues (in brief in (2002) and at great length in his (2000)) that they can never meet the control condition. I agree – I don’t discuss the issue of doxastic voluntarism in this paper, but see my (unpublished a).

Someone believes that  $p$  is probable on evidence  $e_1$ , but that  $\text{not-}p$  is probable on  $e_2$ , both  $e_1$  and  $e_2$  being pieces of evidence she has. She believes, further, that  $\text{not-}p$  is probable on  $e_1$  and  $e_2$ , the total evidence. Could she, nonetheless, go on to believe that  $p$  on the basis of  $e_1$ ?

Hurley claims that she could not. Why not? Hurley writes (1989: 131–2):

Less inclusive probabilistic evidence has no constitutive reason-giving force that could hold out in the face of recognition that it's subsumed by the best probabilistic evidence, which favours the opposite conclusion.

The idea relies a familiar (if somewhat obscure) distinction between practical and epistemic reasons. Outweighed reasons to act in a certain way still count as reasons so to act. Defeated (or “subsumed”) epistemic reasons to believe  $p$ , on the other hand, are no longer reasons to believe  $p$ .<sup>17</sup> A subject can act for an outweighed practical reason because the reason still counts in favor of so acting. But a subject cannot believe  $p$  on the basis of something that she regards as a defeated reason because so regarding it means no longer regarding it as a reason for believing.

But suppose we grant Hurley that the subject couldn't believe  $p$  *on the basis of the defeated* reasons – isn't the real question simply whether the subject could believe  $p$ ?

Hurley grants that a subject could still come to believe  $p$  after concluding that it is unsupported by the total evidence. But, she contends, since the belief can't be based on

---

<sup>17</sup> The phrase “defeated reason”, is, on this view, syncategorematic, for a defeated reason is not a reason. Cf. “decoy duck.”

what the subject regards as a defeated reason, it must arise as the result of either (i) self-deception or (ii) wishful thinking. The belief arises as the result of self-deception if the subject find a way to convince herself that the defeated reason wasn't, after all, defeated – that her overall evidence pointed to p after all. But then we wouldn't have a case of revealed irrationality. The belief arises as the result of wishful thinking if the belief that p is based on a *desire* rather than the defeated reason. But then we wouldn't have a case of revealed irrationality since the subject needn't regard her belief as irrational – she might think the epistemic reasons against believing p were outweighed by practical reasons in favor of believing p.<sup>18</sup> So once a subject believes that the total evidence supports not-p, it is not possible for her to come to believe p *and regard that belief as irrational*: she will either think believing p was supported by the evidence after all or else that it was the thing to believe given her desires.

That, at any rate, is my best attempt at reconstructing Hurley's argument. The lynchpin seems to be the following unsupported assumption:

*No Defeated Basis*: If S regards r as a defeated reason to believe p, then S cannot believe p on the basis of r.

Unfortunately for Hurley, No Defeated Basis is false. For a belief to be based on something is for it to *depend*, psychologically, on that thing: to a first approximation, we

---

<sup>18</sup> Hurley says this is a case of conflict between practical and epistemic reasons, not a conflict between epistemic reasons themselves, which is what epistemic akrasia would require.

can say that if you take away the basis of a belief, then the belief goes with it. But then there is no obvious reason why, due to the vagaries of a subject's psychology, a belief that *p* might be based on what the subject regards as a defeated reason. That reason which she regards as defeated might just be what sustains her belief that *p*.<sup>19</sup>

To see one way this can happen, let's think more carefully about how wishful thinking often works. What wishful thinking typically motivates is focusing one's attention on evidence favorable to the desired hypothesis and not thinking about the evidence against that hypothesis.<sup>20</sup> Even once an assessment of the evidence has led one to conclude that the evidence favors not-*p*, a strong desire for *p* can motivate one to go back and dwell on the evidence for *p* while temporarily not paying attention to one's assessment of the total evidence. It seems perfectly possible for such selective attention to result in the formation of a belief that *p*. What would the belief that *p* be based on? No doubt the belief is partly based on the desire. But it is also partly based on the defeated evidence. To see why, consider what would happen if one lost that evidence: one would be disposed to give up the desired belief. One might, of course, immediately start looking for other evidence to support the belief, but that only reinforces my point: it

---

19 Some theories of the basing relation make believing that *r* is a good reason for *p* necessary for a belief that *p* to be based on *r*. Such a theory of basing would get Hurley the No Defeated Basis assumption. Unfortunately, such theories are implausible. For an argument against such a necessary condition see my (forthcoming).

20 Of course, wishful thinking motivates many other tactics, but I want to keep the present discussion focused.

shows that the evidence was partially sustaining the belief.<sup>21</sup> Apparently, focusing attention on one piece of evidence can give it more psychological force in belief formation (and sustenance) than one's judgment about the import of the total evidence.

Once we appreciate this point, we can see wishful thinking is not the only way that we might believe on the basis of defeated reasons. Perhaps, as a matter of psychological fact, anecdotal evidence tends to grab our attention much more than statistical evidence.<sup>22</sup> Then, when one has some anecdotal evidence pointing one way in conflict with statistical evidence pointing the other way, one might end up focusing attention on the anecdotal evidence and so believing something disconfirmed by the statistics – even if one is aware that the statistical evidence swamps the anecdotal.<sup>23</sup> Or perhaps evidence that confirms pre-existing stereotypes is more psychologically salient than evidence that disconfirms them. Evidence suggesting a causal connection may draw the mind's eye more than evidence suggesting brute correlation. And so on. Even if one properly assesses the evidence in these sorts of cases, one's mind may be repeatedly drawn to the more attention-grabbing subset of the evidence and an irrational belief may be formed on that basis.

---

21 In my "The Problem of the Basing Relation" (forthcoming), I defend the view that a belief that *p* is based on another mental state *q* if and only if one is disposed to revise the belief that *p* when one loses *q*.

22 And there is, of course, empirical evidence that something like this is true.

23 This may be why, for example, so many people are convinced that BLANK vaccinations cause down syndrome.



That's all about forming a belief on the basis of what one has *already* judged to be a defeated reason. But, of course, sometimes one forms a belief on the basis of some evidence and only *later* comes to believe that the total evidence points in a different direction. Here I see no apriori obstacle to supposing that, for one psychological reason or another, one might persist in believing on the original grounds. Why suppose that a reflective belief about the force of the total evidence has some magical ability to necessarily bring one's other beliefs in line with that evidence?

These remarks suggest that the No Defeated Basis principle is false: that we *can* believe on the basis of something we regard as defeated. I don't claim here to have fully defend the possibility of so believing, but I do think I've said enough to shift the burden of proof. Those who want to rely on No Defeated Basis need to mount a defense of it. And that defense will have to be psychologically realistic: it will have to acknowledge the force of the sorts of cognitive biases I discussed above.

So far, then, we have not seen any good reason to reject the possibility of revealed irrationality.

### **Setiya's Dispositional Argument**

In an essay attempting to defend the metaphysical impossibility of believing at will, Setiya suggests that we must endorse the following principle (Setiya 2008: 43):

It is impossible to believe that p or to be confident that p while believing that this degree of confidence or belief is not epistemically justified.

Only with this principle, Setiya suggests, can we understand why it is metaphysically impossible to believe at will. For those of us not antecedently convinced of this metaphysical or conceptual impossibility, however, this is not a particularly persuasive argument.<sup>24</sup> If my rejection of this principle leaves open the conceptual possibility of belief at will, then so be it. In other work (unpublished), I have argued that belief at will is a *psychological* impossibility. And that, it seems to me, is good enough to account for the apparent queerness of doxastic voluntarism.

Setiya does, however, suggest another line of argument for this principle (2008: 43):

In any case, the basic thought is that part of what it is to believe that p – part of what distinguishes believing from other attitudes that might inform behavior, like assuming something, taking it for granted, or accepting it in a context – is the disposition to defend one’s attitude in epistemic terms, as for instance by appeal to evidence that p. Properly characterized, this disposition is inconsistent with believing that one’s attitude is not epistemically justified, which is therefore inconsistent with the belief that p.

This suggestion contains two interesting thoughts: (i) that believing that p implies a disposition to defend p in epistemic terms, and (ii) that this disposition (properly

---

<sup>24</sup> I hold that doxastic voluntarism is *contingently* false, and its contingent falsity is consistent with the existence of revealed irrationality. If, like me, you think being like us can’t believe at will, but you also think we can recognize our own beliefs to be irrational, then you’ll want to see my (unpublished a)

characterized) is inconsistent with believing that one's belief in *p* is unjustified. Let's examine them in turn. Suggestion (i) strikes me as an overintellectualization of the nature of belief. I'm much more inclined to accept the commonsense claim that children and animals have beliefs than I am to accept a speculative metaphysical claim about belief. Now whether animals and children have epistemic concepts is an interesting empirical question. So, to my mind, the plausibility of (i) hangs on research in cognitive ethology and developmental psychology that could go either way.<sup>25</sup> And even if many animals do turn out to have something like epistemic concepts, there is the further question whether any of their social behavior is properly characterized as *defending* their beliefs with those concepts. So, absent empirical evidence that animals (and which ones) exhibit such behavior, I'm hesitant to accept any argument that assumes belief requires a disposition to defend one's position in epistemic terms. Of course, not everyone shares my proclivity to favor commonsense over metaphysical speculation so perhaps for present purposes we can grant Setiya claim (i). Perhaps the claim can be weakened so that it holds only for believers with epistemic concepts.

First, I would like to raise one further worry about the idea that believing *p* implies a disposition to defend *p* in epistemic terms. Let's consider an example from my own doxastic life that I hope will be familiar enough to most readers. For several years in

---

<sup>25</sup> Research with which, unfortunately, I am unfamiliar. There is some evidence, however, that rats are capable of metacognitive reflection on their degree of reliability at various cognitive tasks. So they may possess some sort of working epistemic concepts. For brief discussion, see (Evans & Smith 2012: 36–37).

my mid–20s, I believed that broccoli was a source of complete protein. As it happens, this is true but I had no idea how I picked this belief up. I delighted in relaying this surprising fact to others.<sup>26</sup> For a long time, I never received a challenge to this claim (many fairly bold assertions go unchallenged if said with the right air of authority) and didn't reflect on its grounds. Eventually, though, someone did challenge the claim and a moment's reflection revealed that I found the belief indefensible: I had no idea why I believed that broccoli was source of complete protein and could.<sup>27</sup> My challenger and I looked the nutrition facts up and found that I was right after all.

I maintain that I really did believe that broccoli was a good source of protein. But I also maintain that I was not at all disposed to defend this belief in epistemic terms. For one thing, as soon as the opportunity to mount a defense presented itself, I judged that the belief was indefensible. Further, I did not have any of the following dispositions:

- A disposition to cite some piece of evidence E in support of my belief when challenged (I was aware of no such evidence!)

---

<sup>26</sup> The belief, of course, disposed me to more than merely verbal behavior. I'd sometimes choose to eat broccoli if I thought I hadn't eaten enough protein for the day.

<sup>27</sup> There was the fact that I believed it, which, according to some, is at least some reason to think the belief true. For my part, this doesn't strike me as respectable way of defending a belief.

- A disposition to say, when challenged, that I must have had some good evidence or else I wouldn't have formed the belief (I know, unfortunately, that this isn't true of me)
- A disposition to say, when challenged, that my beliefs are justified until proven innocent (I don't endorse conservatism in epistemology)

What else could a disposition to defend the belief in epistemic terms amount to?

My case generalizes to anyone who rejects conservatism. Whenever such a person believes *p*, but is aware of no grounds for that belief, then I can see no sense in which that person is disposed to defend their belief in epistemic terms.<sup>28</sup> So if Setiya wants to maintain that believing implies this disposition, he must say that people like me can never have beliefs for which they are of no supporting evidence. But, as Descartes took pains to remind us, many of the things we believe are such that we can produce nothing in favor of them when challenged. Whatever rarefied conception of belief rules such a possibility out is not the notion of belief we employ in everyday life.

One might be tempted to say that I (and others in such cases) was at least disposed to *try* to mount an epistemic defense: when challenged, I rooted around in my mind looking for evidence. Perhaps, but it's important to bear in mind that this trying may

---

<sup>28</sup> I'm not speaking of cases where you remember having had some evidence, but you can't remember what that evidence is. (For example, remembering that one read it somewhere.) I'm speaking of cases where you have no idea if you ever had any good evidence.

amount to nothing more than checking whether one can produce grounds in favor of the belief. So, at best, Setiya can defend the following weaker principle:

(i\*) If one believes that  $p$ , then one is disposed to check whether one has grounds for  $p$  when challenged.<sup>29</sup>

As it happens, this disposition is not inconsistent with believing one's belief that  $p$  to be unjustified. To see why, let us first consider why Setiya likely takes believing one's belief unjustified to be inconsistent with being disposed to defend that belief in epistemic terms. A characteristic of believing  $p$  is being disposed to assert  $p$  in relevant contexts. So if one believes one's belief unjustified, one will be disposed to assert that one's belief is unjustified in relevant contexts. One relevant context is when one's first-order belief is challenged. But asserting one's belief is unjustified is inconsistent with defending one's belief in epistemic terms. So, on the assumption that one cannot be disposed to  $\phi$  when  $C$  and to  $\psi$  when  $C$  if  $\phi$  and  $\psi$  are inconsistent,<sup>30</sup> one cannot be disposed to assert that one's belief is unjustified when challenged and disposed to defend one's belief when challenged. Hence, one cannot believe that one's belief is unjustified while being disposed to defend that belief in epistemic terms.

---

<sup>29</sup> Even this weakened principle I wonder about. What about beliefs that are very important to a person's self-conception, where finding out that they're false or ungrounded could be extremely unpleasant? Might we not sometimes be disposed to avoid checking whether those beliefs are grounded? But perhaps in such cases, we're disposed to (unjustifiedly) respond to challenges by saying things like "I just know it," which is a sort of epistemic defense. So Setiya might say one is either disposed to check for grounds or to buntly defend one's belief.

<sup>30</sup> This assumption is dubious, but I pass over it for now. It will receive further scrutiny in due course.

That argument is too quick, but we need not slow it down, for we have rejected the claim that believing  $p$  implies being disposed to defend  $p$  in epistemic terms. What we are granting is that believing  $p$  implies being disposed to check for grounds for  $p$ . So the question that now confronts us is whether this disposition is inconsistent with believing that one's belief that  $p$  is unjustified. We should check, then, whether that disposition is inconsistent with the following disposition: the disposition to assert that one's belief that  $p$  is unjustified. On the face of it, these dispositions are not inconsistent. For one thing, one can assert that one's belief is unjustified while checking to see whether one has grounds (just to make sure, as it were). So it is not available to Setiya to say that these dispositions are inconsistent on the grounds that their manifestation is inconsistent. One might be tempted to suggest that believing that  $p$  implies being disposed not to double-check whether one has evidence inconsistent with  $p$ . Then believing one's belief that  $p$  to be unjustified would imply being disposed not to double-check whether one has evidence for  $p$  (for this would amount to looking for evidence inconsistent with one's meta-belief). But this isn't true: believing  $p$  doesn't imply being closed to reevaluating one's grounds for  $p$ .

More to the point. Suppose, for the sake of argument, that I believe  $p$ . Suppose I also believe that my belief that  $p$  is unjustified. Suppose someone finds out that I believe  $p$  and challenges me. Here's what will probably happen: I'll feel the temptation to check for grounds for my belief (I'm so disposed), but I'll suppress the urge because I believe

that I won't find any. Though I'm still disposed to check for grounds for p, my belief that p is unjustified *masks* that disposition. As far as I can tell, this is a perfectly possible dispositional profile. What will I say to the person who challenged me on p? "I know I (epistemically) shouldn't believe p, but I can't shake it" (or "it makes me feel good to believe").

But suppose I'm wrong that believing p implies at most a disposition to check for grounds: suppose it implies, after all, a disposition to defend p in epistemic terms. That is not inconsistent with also believing that one's belief that p is irrational. For this latter belief could mask one's disposition to defend p in epistemic terms, while still leaving one with the relevant disposition. Whatever sorts of counterfactuals are entailed by dispositions, it's widely agreed that they're compatible with this sort of masking.

So I think that Setiya's argument fails. He has identified nothing about the dispositional profile of believing p that is inconsistent with believing that that belief is unjustified. What's more, this discussion has begun to shed some light on how such a conflicted doxastic state might be possible. I would like to turn now to elucidating this possibility. Common sense and ordinary language tell us that such doxastic conflict is possible. We have examined several prominent philosophical arguments against the possibility and found them wanting. We should conclude, therefore, that the *prima facie* case is intact. Still, this sort of conflict *is* puzzling: we would do well to gain a



philosophical understanding of how it might arise. I think things become considerably less puzzling when we distinguish clearly between occurrent judgment and standing belief.

### **Occurrent Judgment vs. Standing Belief**

Two distinctions are often drawn between types of belief: occurrent vs. non-occurrent beliefs and dispositional beliefs vs. dispositions to believe. The second distinction is intuitive enough: many propositions I've never before considered are so obvious — or follow so obviously from other things I believe — that I would immediately believe them upon considering them. We want to say that though I don't yet *believe* them (in some sense), I am now *disposed to believe* them. (This is not the distinction between conscious and unconscious beliefs: things I *now believe* concern propositions I've never consciously considered or endorsed.) Of course, to call the things I already believe *dispositional beliefs* is rather contentious: not everyone agrees that belief is properly understood as a dispositional state. It's better, I think, to adopt Lycan's terminology and distinguish between *explicit beliefs* and *tacit beliefs* (Lycan 1988).

While that distinction is relatively clear (even if a theoretical account remains elusive), the distinction between occurrent and dispositional beliefs is obscure. Audi, who has had much to say on the topic, observes that, “[occurrent belief is] not an occurrence that *is* a belief (since beliefs are not events in the relevant sense) but, roughly, belief

whose propositional object is in some way occurring to one,” (1994: 420). Of course, there are many ways a proposition can occur to one: one can consider it, or feel doubts about it, feel afraid about it, etc. All of these events count as occurrent beliefs on Audi’s usage. But that is not what others have in mind. Stich writes (Stich 1978: 504):

[I]f we ask a subject whether *p* is the case, he will generally have a certain sort of characteristic experience which, as best I can discover, has no standard description in English. Some philosophers have labelled the experience ‘having the occurrent belief that *p*.’ ... [I]n typical cases of belief a subject will have a certain sort of characteristic conscious experience when his attention is suitably directed to the content of the belief.

As this passage from Stich suggests, the experience is not to be identified with the belief, but is caused by it.

On either understanding, the term “occurrent belief” is best thought of as syncategorematic: occurrent beliefs are not beliefs but occurrences (thoughts, experiences, or whatever) *associated with* or *caused by* beliefs. Similarly, “non-occurrent belief” does not, on this usage, characterize a type of belief: it simply marks that the subject is not, at the moment, experiencing the relevant occurrence.

Audi’s conception of occurrent belief strikes me as idiosyncratic; at any rate, it is the phenomenon Stich alludes to that interests me here. But since I think the phrase “occurrent belief” tends to cause more confusion than elucidation, I prefer to call the occurrence in question a *judgment* (Lycan 1988: 55). One might think of judging as the

“interior analogue” of assertion<sup>31</sup> – silent assertion to oneself. The paradigmatic case is when the question whether *p* occurs to one, and one thinks, “Yes, *p*”, or “No, not-*p*” (the first is judging that *p*, the second, judging that not-*p*) in that special way that is as difficult to describe as it is familiar. In the paradigmatic cases, judging that *p* has a distinctive phenomenology: it alters what things are presently like for one (Silins 2012).

So there is the occurrence of a judgment on the one hand, and the existing of a standing belief on the other. What is the relation between these two attitudes? Plausibly, a belief that *p* typically *causes* a judgment that *p* in the relevant circumstances. That is why, typically, when we believe *p* and someone asks us whether *p*, we find ourselves judging that *p*. This makes it plausible that judgments are a source of evidence about what we believe – if you find yourself judging that *p*, that’s reason to think you believe that *p*.<sup>32</sup>

But it doesn’t seem that there is anything deeper than this causal and epistemological connection. In particular, I see no reason to suppose that believing *p* entails judging *p* in the right circumstances, or that judging *p* entails believing that *p*. Even if it is part of the nature of belief to dispose one to judge that *p*, surely something else could mask that disposition from manifesting itself. Suppressed beliefs, for example,

---

31 To borrow a suggestive phrase from Williamson (2000).

32 For a defense of the view that judgments provide a source of non-inferential justification for beliefs about what we believe, see (Silins 2012). For dissent that we learn about our mental states via conscious experiences that are distinct from those states, see Shoemaker (1996, ch. 2). As far as I can tell, nothing I say here commits me in either direction: even if beliefs are inherently self-intimating, that doesn’t call into question the existence of current judgments that are not beliefs. Indeed, the existence of such judgments seems essential to reasoning about evidence that contradicts what we already believe.

seem to work something like this. Further, if we think of judgment as a conscious occurrence individuated by its phenomenology, there is no reason to suppose that one could undergo that phenomenology in the absence of a belief. Silins provides the following example (2012: 16):

[Y]ou judge that your flight leaves at noon, and then realize that you do not and did not believe that your flight leaves at noon. In such cases, your judgment that p is a kind of performance error which fails to reflect an underlying belief---“what was I thinking?”, you might go on to say. You “blurted out” that p, either in speech or in merely in thought, consciously endorsing the proposition that p, yet failing to have a standing belief that p.

Consider an (admittedly limited) analogy with perception. Typically, if one seems to see p, that is because the fact that p caused one to undergo that experience. But it is perfectly possible to seem to see p when p is false (and so didn't cause one's experience). Of course, one might not be able to *see* p unless the fact that p caused one's experience, and so perhaps there is an analogous use of the term judgment: one does not really judge that p, on this usage, unless a belief that p caused the experience. But that is a terminological dispute. There will still be a phenomenologically indistinguishable state of *seeming* to judge that p. It is that type of experience, on my usage, that counts as a judgment that p. The analogy breaks down, of course, because judging that p typically causes one to believe that p (if one didn't already). But, again, there is no necessity to this causal connection: things can interfere so that one doesn't end up believing the proposition judged to be true.

We are now in a position to see more clearly what is wrong with Adler's fourth premise: that one cannot disbelieve the conclusions of reasoning. To draw a conclusion in a piece of reasoning is to make a judgment. But that an event of judging has occurred does not entail that any standing state of belief exists.

Further, we can now see how someone might get into a state of revealed irrationality. Let us return to our skeptics, and retrace their reasoning, but this time being careful to distinguish between belief and judgment:

(T) The skeptic forms the belief that she believes that p by judging that p (and this judgment is caused by her belief that p).

(T+1) The skeptic judges that it is irrational to believe p (she derives this from her skeptical thesis) and rescinds her judgment that p — judging instead that p is unsupported by her evidence.

(T+2) So the skeptic judges that she irrationally believed, at t, that p.

(T+2) The skeptic continues to believe that p.

(T+2) The skeptic continues to believe *inferentially*, at t+2, that she believes p (and is unlikely to stop anytime soon), on the basis of beliefs about her behavior, etc.

(T+3) The skeptic *judges, of her own belief that p, that* it is irrational and *that* she continues to believe it.

(T+4) This judgment causes the skeptic to believe, of her own belief that p, that it is irrational and that she continues to believe it.

It is possible for the skeptic to rescind her judgment that p at (T+1), while continuing to believe p, because there is no necessary connection between belief and judgment. Though her belief that p may dispose her to judge that p, her reminding herself of her commitment to skepticism blocks this disposition. But the skeptic can still judge that she believes that p at (T+2) because occurrent judgments are not the only way to form beliefs about what we believe: we can also form such beliefs by reflecting on our behavior and tendencies. Once she's gotten this far, the skeptic is then able to judge, and come to believe, that her own belief is irrational.

I think clearly distinguishing between judgment and belief makes it impossible to argue that someone couldn't end up just like the skeptic at (T+3): judging that her own belief is irrational. And perhaps that is all we need to make sense of the phenomenon of revealed irrationality. Perhaps there is some further argument about the nature of belief that prevents the transition from (T+3) to (T+4) — but I can't for the life of me see how it would go. So it is time, finally to leave behind this old bugbear: revealed irrationality is possible; and the examples with which we began show that it is actual.

## CHAPTER 2: THE DISPOSITIONALIST GAMBIT

Questions about what sort of control we have over our beliefs are important for the “ethics of belief.” So-called *doxastic voluntarists* are united in holding that, at least sometimes, we have fairly “direct” voluntary control over what we believe.

Unfortunately, just what counts as *direct* or *voluntary* is part of what is in contention.

There are nearly as many distinct doxastic voluntarist theses as there are doxastic voluntarists. Similarly, there are many distinct strategies for defending such theses.

Doxastic voluntarism has become, to borrow a colorful phrase of Bonjour’s, “a philosophical hydra: difficult to come to grips with and seemingly impossible to kill” (1978, p. 1).

My aim in this paper is modest: I identify one strand of voluntarism and examine a recent strategy for its defense – what I call *the dispositionalist gambit*. The first part of the paper is devoted to getting a sufficiently clear and interesting voluntarist thesis in view. I then turn my sights on two implementations of the dispositionalist gambit – from Ginet (2001) and Frankish (2007) – and argue that neither succeeds. I conclude by tentatively suggesting an argument that voluntarism is (contingently) false.

### **What is Doxastic Voluntarism?**

In recommending belief in God as the prudent option, Pascal faced a problem: those with genuine doubts about the existence of God would find it difficult to suppress their doubts and become believers, even if they wanted to. Pascal's solution was to recommend a program: the doubters should immerse themselves in Christian life; they should go to church, pray, read the bible, speak with other Christians about their religious experiences, and so on. Further, they should strive not to think about evidence against God's existence but focus instead on evidence in favor of God's existence. Pascal was confident that such a program would eventually terminate in converting even radical atheists into firm believers. This is not controversial: one can often decide to submit oneself to a program of indoctrination with a reasonable expectation of success.

Now this already raises a problem for formulating doxastic voluntarism. On the one hand, there is a temptation to say that because believing is involuntary, Pascal could not *simply* recommend that atheists believe in God. On the other hand, we're tempted to say that because we do have *some* control over what we believe, atheists convinced by Pascal need not sit around idly *wishing* they believed in God – they can *do* something about it.

This has led many to distinguish between *direct* and *indirect* voluntary control over our beliefs.<sup>33</sup> The sort of control we uncontroversially have over our beliefs is at best

---

<sup>33</sup> This distinction pervades the literature on doxastic voluntarism. See, for example, (Feldman 2008), (Bennett 1990), and (Winters 1979).



indirect: there are things we can do that, if things go right, can bring about a desired belief. This can be contrasted with, say, raising my arm, which is something I can do *directly*. With this distinction in hand, we might characterize the *interesting* doxastic voluntarist thesis as the claim that we sometimes have direct voluntary control over what we believe.

Of course, this notion of direct control needs elucidation. Generalizing from the examples, the obvious elucidation comes from the concept of *basic action*: a basic action is one that is not composed of further actions. Raising my hand is a basic action; carrying out an elaborate process of Pascalian indoctrination is not. With this conception of directness, the interesting voluntarist thesis would be that we can come to believe something as the result of a basic action.<sup>34</sup> This is, roughly, the thesis that Bernard Williams has famously attacked.

This thesis has as good a right as any to the title of “doxastic voluntarism” and it is indeed interesting. But it is too strong: there are weaker theses that are just as interesting. Notice that on this notion of directness, we lack direct voluntary control over whether our shoes are tied – tying one’s shoes is a non-basic action. Similarly, we don’t have direct voluntary control over making omelettes,<sup>35</sup> packing our bags, or driving to

---

<sup>34</sup> Or, perhaps better, that there is a basic action of belief formation. Indeed, this is precisely how Frankish characterizes his own voluntarist position.

<sup>35</sup> Though making an omelette may well be a basic action for Jacques Pepin.

work. But many of us want to maintain that forming beliefs is not much like tying one's shoes, making an omelette, packing one's bag, or driving to work!<sup>36</sup> So even if there is no basic act of belief formation, it would still be a striking doxastic voluntarism thesis that claims we have the sort of control over our beliefs that we have over these other non-basic actions.

What sort of control is this? Notice that (typically) when we tie our shoes, this is something we intended to do.<sup>37</sup> When we form a particular belief, on the other hand, this is not (typically) something we had intended. Perhaps we could understand voluntarism as the thesis that we sometimes intend to form and revise beliefs. But this won't do. For one, delusional people can intend to bring about all sorts of things over which they have no control (they can intend to bring about rain, or to influence the outcome of the Super Bowl). And, more to the point, forming a particular belief is precisely what the atheist following Pascal's advice intends. Instead, I think we should focus on the fact that is not clear that belief formation is an action at all – basic or otherwise. As Ryle has pointed out, this marks one important contrast we have in mind when we distinguish between the voluntary and the involuntary:

---

36 I am not the first to have noticed this: Heil (1983: 358), Hieronymi (2006: 48–9), and Setiya (2008: 41) all make similar observations.

37 Whether we had a conscious intention to tie our shoes is another matter – generally speaking, we don't have such conscious intentions.

What is involuntary, in [one] use, is not describable as an act. Being carried out to sea, or being called up, is something that happens to a person, not something which he does... So sometimes the question ‘Voluntary or involuntary?’ means ‘Did the person do it, or was it done to him?’ (Ryle, 2009, 60)

Let us pursue this thought. It is of course true that the atheist following Pascal’s advice, if successful, forms a belief as the result of some intentional actions of his. Similarly, my shoes are tied as the result of an intentional action of mine. But there may still be a difference: my shoes are tied as the result of a specific (non-basic) action: that of *tying my shoes*. My shoes are not merely tied as the result of a series of basic actions. That series of basic actions *constitutes* tying my shoes. To see what I mean, consider a British naval officer who lost the use of his arms in the Battle of Trafalgar. Now, when he needs his shoes tied, he rings a bell and instructs his manservant to do it for him. We can assume that this series of actions reliably results in his shoes becoming tied. But when he performs these actions, he is not tying his shoes – if he could tie his shoes, he wouldn’t need the manservant! These actions constitute *requesting that his manservant tie his shoes* but not *tying his shoes*.

Further evidence for this distinction results in the sort of advice that we can helpfully offer. If you keep tripping over your laces, I can helpfully advise you to tie your shoes. If our British naval officer keeps tripping over his laces, it would be cruel to advise him to tie his shoes. (We should instead tell him to call his man, or perhaps to hire a less

negligent servant.) So, in general, there is a distinction between a series of actions merely resulting in  $\phi$  and a series of actions constituting  $\phi$ -ing.<sup>38</sup>

It is worth noting that this distinction is orthogonal to Williams's distinction between action that is *intentional* under a description and action that is *unintentional* under a description. When a series of basic actions constitutes  $\phi$ -ing, it is still open whether the  $\phi$ -ing was intentional. A careless driver who makes a right turn into a bicyclist has performed a series of non-basic actions that can be described as *hitting a bicyclist*. Under that description, the action is unintentional – but it is, nonetheless, an action performed by the driver. The actions of the driver don't merely result in a bicyclist being hit, they constitute hitting a bicyclist. In contrast, our British naval officer's actions are not appropriately described as *tying his shoes*. So there's no further question whether he *intentionally* tied his shoes.

Returning to our Pascalian atheist, it does not seem right to say that his program of indoctrination *constitutes* forming the belief that God exists. He hopes that it results in the formation of that belief, but he could not, on his way to church, accurately report on what he was doing by saying "I am forming the belief that God exists."<sup>39</sup> And, Pascal's advice is consistent with this diagnosis: Pascal does not advise the atheist to form the

---

38 Interestingly, a series of actions can constitute  $\phi$ -ing even if they don't result in  $\phi$ . I might be shot dead in the middle of tying my shoes. It's still true that I was tying my shoes.

39 He could say with perfect propriety, "I am trying to form the belief that God exists." In general, when no series of basic actions available to one constitutes  $\phi$ -ing, there will still be some series of basic actions that constitutes trying to  $\phi$ .

belief that God exists; he advises him to start going to church and stop dwelling on arguments against God's existence.

I propose, then, that one interesting strand of voluntarism concerns not whether forming a belief is something we have "direct control" over, but whether it is something we *do* or something that merely *happens to us* (allowing, what is uncontroversial, that it sometimes happens to us as a result of other things we do). So this paper will concern itself with the following thesis:

*Doxastic Voluntarism*: There are intentional actions an agent can sometimes perform, such that in performing them, he is forming (or revising) a belief.<sup>40</sup>

I don't claim that this is the only thesis that has a right to the title "doxastic voluntarism," only that it is one that is reasonably tractable and interesting.

Now, my aim in this paper is to evaluate a recent strategy for defending voluntarism: the dispositionalist gambit. If successful, the gambit would establish doxastic voluntarism as I've defined it here.<sup>41</sup> Whether the gambit might be employed to establish a stronger version of voluntarism is not something I'll explore. Similarly, I will

---

40 Audi (2001: 94) formulates voluntarism similarly as the question whether belief-formation is an action type – but he (seems to) restrict the thesis to its being a *basic* action-type. Audi points out an interesting variant of the thesis: the claim that there are acts of belief *revision*. I leave the examination of that thesis for another day, though it seems to me that the considerations I raise create problems for this variant as well.

41 Though the authors I discuss do not formulate voluntarism as I do: their question is whether one can decide to believe. But they identify an action that they claim constitutes forming a belief,

not explore whether the anti-voluntarist argument I offer tells against different versions of voluntarism.<sup>42</sup>

### **The Dispositionalist Gambit**

Consider the following suggestive remarks from Harman:

What is it to believe something? At least two things are involved, both of which are things one can decide to do. Beliefs serve as the basis for other acts one may undertake, and beliefs end one's inquiry into an issue. One can certainly decide to accept a given proposition as a basis for action; that is one can decide to end one's inquiry into an issue and hold to a certain conclusion. (2005, p. 95)

Here, Harman suggests that one can decide to believe something because two of the things involved in believing are things that one can decide to do. Of course, all by itself, this hardly constitutes an argument that one can decide to believe. For there may be other things required of belief that one *cannot* decide to do.<sup>43</sup> Nevertheless, Harman presents a strategy for defending voluntarism: identify what is required to believe that p, then inquire into whether one can decide to do those things that are required.

Of course, I have not defined doxastic voluntarism as the thesis that one can, sometimes, decide to believe that p. But Harman's strategy can be adapted to my favored formulation. Harman has defined the state of belief that (partially, anyway) in terms of

---

<sup>42</sup> Some versions of voluntarism claim that a belief state is voluntary even if there is no voluntary action of forming (or revising) beliefs (Steup 2008; Weatherson 200). Whether and how the argument offered here applies to such views is a delicate issue I don't take up here.

<sup>43</sup> Indeed, I will suggest later that Harman himself identifies something else required of belief that one cannot decide to do.

two other states: (i) being disposed to take p as a basis of action, and (ii) being disposed to inquire no further into the question whether p. We can then ask whether there are intentional actions one can perform such that, in performing them, one is disposing oneself to take p as a basis of action or disposing oneself to inquire to further into whether p. Or, more simply, whether there are acts of adopting the relevant dispositions. If there are, and Harman is right that (i) and (ii) are sufficient for belief, then we are on the road to identifying possible acts of intentional belief formation.<sup>44</sup>

Harman did not intend to mount a full defense of doxastic voluntarism along these lines. But other authors recently have. The dispositionalist gambit depends on two plausible theses:

*Dispositionalism About Belief:* In order to believe that p, it is sufficient that one has certain dispositions, BD.

*Intentional Dispositions:* Some dispositions, ID, we can adopt just by forming an intention to do so.<sup>45</sup>

Intentional dispositions are voluntary, in my sense. Forming an intention to be disposed to  $\phi$  can itself be an intentional action – it is something one can intentionally do. If such

---

<sup>44</sup> Plausibly, if one can intentionally act so as to get oneself into states (i) and (ii), then one can act so as to get oneself out of states (i) and (ii). In that case, we might also be on the road to identifying possible intentional acts of belief revision.

<sup>45</sup> This may be a stronger principle than the dispositionalist needs. All she really needs is that some dispositions are such that there are acts of adopting them. But it seems natural to me to suppose that the way the sort of dispositions in question are adopted is by the formation of the relevant intention, so I proceed in these terms.

dispositions can be adopted by this intentional action, it seems to right to say that in performing this voluntary action, one is forming the relevant disposition. So, if the dispositions involved in believing that p are intentional dispositions – that is, if the BD are ID – then doxastic voluntarism is true as I am understanding it. So the *dispositionalist gambit* defends doxastic voluntarism by holding both Dispositionalism About Belief and Intentional Dispositions and by defending the following the thesis:

*Dispositionalist Doxastic Voluntarism*: The BD are ID.

In what follows, we will examine two implementations of the dispositionalist gambit. But first, I want to make a few remarks about Dispositionalism About Belief and Intentional Dispositions, for neither are uncontroversial.

Dispositionalism about belief, in fact, is highly controversial. Dispositionalism about belief can be contrasted with *representationalism*, which holds that in order to believe that p, it is necessary that one contain an internal representation of p in one's "belief box." Representationalism about belief is dominant in the philosophy of mind<sup>46</sup>, though dispositionalism is common enough among epistemologists.<sup>47</sup> The distinction between dispositionalism and representationalism is complicated by the fact having an internal representation "in the belief box" is typically understood in functionalist terms,

---

46 See, e.g., (Fodor 1985), (Dretske 1988), and (Cummins 199).

47 See, for example, (Audi 1994) or (Williamson 2000: 47). One of the more prominent contemporary philosophers of mind who defends this sort of view is Eric Schwitzgebel; see, for example, his (2002).



and dispositionalism can be understood as a limiting case of functionalism (where the functional role isn't picky about about inputs, only outputs). Nonetheless, it seems that the dispositionalist gambit is committed to denying certain versions of representationalism. I won't question that commitment here, but it is worth flagging.

The Intentional Dispositions thesis is less controversial, if less clear. The idea rests, I think, on some plausible intuitions. What sorts of dispositions can we plausibly adopt just by intending to? I might "decide" tonight that, from now on, I shall wake up at six in the morning. Have I – just like that – disposed myself to wake up at six in the morning? Probably not. Unless I change my alarm clock, I will continue to wake up at 7:30. Upon resolving to start waking up at six, of course, I may very well decide to set my alarm to six. The result will be that I wake up at six every morning. But *I* won't be disposed to wake up at six. My dispositions are, on a plausible view of dispositions, a function of my intrinsic properties. After changing my alarm clock, I am still disposed to wake up at 7:30 – that's precisely why I need the alarm clock!<sup>48</sup> In time, I may get used to it and become disposed to wake up at six o'clock on my own.

Consider a different case. Suppose I decide that, from now on, I won't work on weekends. Have I – just like that – disposed myself to not work on weekends? It seems plausible to say that I have. Unless I am a compulsive "workaholic," forming an intention

---

<sup>48</sup> Compare: after packing a glass in styrofoam, it will no longer break when dropped; but the glass itself, of course, is still fragile – still disposed to break when dropped.

not to work on weekends changes my dispositions. In firmly resolving not to work on weekends, I adopt a disposition to not work on weekends.

Many philosophers I know are workaholics, so this may be a bad example. But consider this: suppose I eat lunch everyday at a nearby Thai restaurant and always order the same thing. Tired of my predicatbility, I decide that from now on I won't order a menu item unless it his been at least five days since I last ordered it. That decision (or intention) changes my dispositions. I'm now disposed not to order the same thing everyday, just like that.

One might wonder why the intention changes my dispositions in the one case and not the others. I'm afraid I don't have anything particularly informative to say on the matter. The difference *seems* to hang on the fact that I don't have the sort of control over when I wake up that I do over whether I work on the weekend. But spelling out just what sort of control this invokes is a formiddable task (certainly I have some sort of control over when I wake up – that's what the alarm clock buys me). Still, the intuitive difference is clear and that's probably good enough for the dispositional gambit. At any rate, it's certainly good enough for my purposes since I will argue that the strategy fails for other reasons.

### **Ginet's Implementation of the Gambit**

Carl Ginet has employed just this strategy in arguing for doxastic voluntarism. His account begins with the notion of *counting on a proposition*.<sup>49</sup> Counting on a proposition *p* is something we sometimes do when deciding on a course of action. It amounts to choosing a course of action that you believe optimal iff *p*, and also to being dismissive about the possibility that not-*p* (Ginet 2001: 65). According to Ginet, we can decide to count on *p*. That involves (i) deciding to perform the action that is optimal iff *p*, which is something we can clearly do, and (ii) deciding to be dismissive of the possibility of not-*p*, which is not so clearly something we can decide to do.

Part of being dismissive of not-*p* is not preparing oneself for that possibility (66). That is surely something one can do. But Ginet also holds – and needs to hold, I think – that being dismissive of not-*p* involves not giving that possibility much thought and of being surprised if one discovers that not-*p* is in fact true. Now one can decide not to think much about the possibility of not-*p* in the sense that one can suppress or shift away attention from not-*p*. But one cannot, of course, prevent those thoughts from creeping before one’s mind in the first place. And even if one can suppress any conscious awareness of such thoughts, they can no doubt lurk in the background and cause all sorts of stress reactions about, say, the burner one worries one might have left on back in the

---

49 For a classic analysis of acting on a proposition, see (Chisholm 1961). Chisholm there distinguishes four senses of the phrase “act on a hypothesis *h*.” The sense most relevant to present discussion Chisholm defines as follows: “S acts on a hypothesis *h* =df. in so acting, S relies upon *h*” (5). Chisholm then defines relying on *h* as not providing for the possibility of not-*h*.

house. Further, it is hard to see how one could decide to be surprised were one to discover that not-p.

If one is aware of good reasons to think that not-p is a live possibility, and that one's plans hinge on p, deciding to ignore the possibility may be like deciding to wake up at six o'clock: something one can't pull off just like that. Worrying thoughts of not-p will likely creep before one's mind and disturb one's complacency.

Couldn't Ginet just abandon these problematic aspects of being dismissive? Couldn't he instead work with the simple notion of not preparing for the possibility of not-p? It seems to me that he cannot, for he wants to parlay his account of deciding to count on a proposition into an account of deciding to believe. People who really believe p don't spend much time consciously worrying about the possibility that not-p. Nor do they harbor unconscious worries about that possibility that trigger stress responses (setting aside various pathological cases). They tend to be surprised if they discover that not-p. If that's right, that Ginet needs to show that we can decide not to worry about not-p and to be surprised should we discover that not-p. Now he does admit that being dismissive of not-p can be challenging and effortful: it may require active suppression of thoughts of not-p and focused attention on the evidence for p. Perhaps Ginet can hope to argue that in some cases, this tactic will eradicate any unconscious worries and will dispose one to be

surprised should one discover that not-p. Perhaps cognitive dissonance mechanisms could be invoked to explain how this goes.

I want to set aside this worry for now and get the rest of Ginet's proposal on the table. So let us grant that we can decide to be dismissive of the possibility that not-p. Then, since we can also decide to act in a way we believe optimal iff p, we can decide to count on a proposition. Now counting on p in choosing a particular course of action is a one-off event that falls short of being in the dispositional state of believing p. But Ginet contends that in deciding to count on a proposition, one "simultaneously adopts and manifests" a disposition to count on p in situations relevantly like the present situation (67). And having such a disposition, according to Ginet, is one way of believing that p. Ginet denies that a disposition to count on p in some situations is necessary for belief – there may be other ways of believing a proposition. And now Ginet's implementation of the dispositionalist gambit falls into place: since one can decide to adopt a disposition to count on p in various situations, and since being so disposed means believing p, one can decide to adopt the belief that p.<sup>50</sup>

Let us descend from abstraction and consider one of Ginet's examples of deciding to believe:

---

<sup>50</sup> Holton (1994) considers a similar strategy for defending voluntarism involving trust: we can decide to trust – in so deciding, do we decide to believe? Holton says "no" because trust need not involve belief.

We have started on a trip by car, and 50 miles from home my wife asks me if I locked the front door. I seem to remember that I did, but I don't have a clear, detailed, confident memory impression of locking that door (and I am aware that my unclear, unconfident memory impressions have sometimes been mistaken). But, given the great inconvenience of turning back to make sure and the undesirability of worrying about it while continuing on, I decide to continue on and believe that I did lock it. (2001, p. 64)

As Ginet would have us understand the case, in deciding to continue on, he has decided to count on the proposition that he locked the door. And in doing that, he has adopted and manifested a disposition to count on that proposition in similar circumstances. If, for example, he calls his mother to let her know that they're on their way and she asks if he remembered to lock the front door, he'll say "Yes, Mother" without giving it a second thought. But given his lack of a clear, detailed, confident memory impression he has certainly not disposed himself to count on having locked the door in *all* circumstances. For example, if his neighbor were to call and let him know about a string of burglaries, he very likely would decide *not* to count on the proposition and ask the neighbor to make sure his front door was locked.

So does Ginet, after deciding to continue on, believe that he locked the door?

Ginet says the belief is context dependent: in some contexts, like the context of deciding whether to continue on, he believes it, while in other contexts, like the context of asking the neighbor to check the door, he doesn't believe it. The idea is not that he would change what he believes were he to learn about a burglary. For he's already simultaneously disposed to count on the proposition in some contexts but not in others. We cannot simply

ascribe belief or disbelief to Ginet (according to Ginet): all we can say is that relative to one context he does believe that the door is locked, but that relative to another context, he does not believe this.<sup>51</sup>

But this is a strange result. If there are contexts in which Ginet wouldn't count on the proposition that he locked his front door – or, at least, if there are sufficiently many not-too-strange such contexts – then isn't he too unsure of whether he locked his door to count as believing it? That would seem to be the more natural thing to say. Indeed, suppose that the *only* contexts in which Ginet is willing to count on the proposition are contexts in which he's deciding to continue on with a car trip. On Ginet's account, he believes in such contexts that his door is locked, but in all others does not believe this. But the right thing to say about such a scenario is that *Ginet doesn't believe that his door is locked*; rather, though he's unsure whether it's locked, he's willing to risk it for the sake of continuing with car trips.

Or consider another example: I grant a proposition for the sake of argument. I might be willing to count on p, though I currently believe not-p, for the sake of seeing where someone's argument goes, but in no other contexts. According to Ginet, I believe p in that argumentative context, but in all other contexts I disbelieve p. That sounds wrong.

---

<sup>51</sup> This is not intended, I don't think, as a commitment to linguistic context-sensitivity for determining the semantic content of "believes that" locutions. The idea is rather that the differing dispositions of the agent constitute different situation-relative beliefs. They are, in effect, contradictory beliefs, though they don't conflict because they are circumscribed to different situations.

I don't believe *p* in the argumentative contexts – I have just shown myself willing to *assume* that *p* for the sake of argument. Being willing to assume *p* in various contexts is one thing; actually believing it is another. We don't need a theory of the difference between belief and contextual assumption to intuitively recognize it. (Though I will offer an account of this distinction in due course.)

Without good, independent reason for positing massively contextual beliefs, we should reject Ginet's implementation of the dispositionalist gambit. Ginet has not, I think, given us reason to think there are intentional acts of belief formation. Rather, all he has shown is that people are sometimes willing to assume, in some contexts, propositions they don't believe to be true. Beliefs don't seem to be contextually limited in the way that assumptions are. This is not, however, an inherent weakness of the dispositionalist gambit: as we shall see below, Keith Frankish runs the gambit without making himself vulnerable to this criticism.

### **Frankish's Implementation of the Gambit**

Keith Frankish has recently called our attention to a disposition to use *p* as a premise in theoretical and practical reasoning in an *open-ended* range of contexts. Using *p* as a premise in a bit of reasoning is something we can do, whether or not we believe *p*. In a bit of reasoning where *p* is relevant, we can decide whether to use it as a premise. But then it is plausible that we can form an intention, from here on out, to use *p* as a premise



in any reasoning where it may be relevant. Forming such an intention, we may expect, amounts to adopting a disposition to take *p* as a premise in any context.<sup>52</sup> And having this disposition, according to Frankish, is what it is to believe that *p*.<sup>53</sup> So if Frankish is right about this, forming such an intention amounts to forming the belief that *p*.

Now one major advantage Frankish's account enjoys over Ginet's is that it doesn't make beliefs massively contextual. Frankish doesn't say that for believing *p*, it is sufficient to use *p* as a premise in *some* contexts; rather, one must be disposed to take *p* as a premise in an open-ended range contexts where it might be relevant. Someone who assumes *p* merely for the duration of an argument, or to have peace of mind during a road trip, does not count as a believer of *p*, according to Frankish – there are many contexts in which such a person would not use *p* as a premise. Frankish claims another benefit for his account: he says it explains why we cannot consciously decide to believe *p* unless we have some evidence for *p*, which he takes to be a datum (2007, p. 532). Why not? Because, Frankish says, one context in which *p* is relevant is a context in which all one cares about is using truths as premises (Frankish calls these *TCP* contexts). Frankish considers it a conceptual truth that believing *p* means being disposed to use *p* as a

---

52 Of critical importance to Frankish's account are contexts in which the only thing one cares about is using true premises. That believing *p* means, in part, being disposed to take it as a premise in such contexts allows Frankish a clever – though ultimately flawed, I think – explanation of why we can only voluntarily adopt beliefs for which we have some epistemic support (Frankish 2007: 536–8).

53 One can find a similar strand of thought in (Van Fraassen 1984): “epistemic judgments” are commitments we take on to reason in certain ways. Van Fraassen's concern, of course, is with credences rather than all out belief and so there won't be a question of using *p* as a premise; rather, the commitment is to use a certain probability when reasoning about *p*.

premise in TCP contexts. But it would be foolish to decide to take p as a premise in a TCP context unless one thought one had evidence for p. So this is not something one can consciously do (2007, pp. 537–8).<sup>54</sup>

That is how Frankish employs the dispositionalist gambit. We should first examine more closely Frankish’s assumption that we can decide, from here on out, to use any old proposition in our conscious reasoning. Let’s begin with practical reasoning. So we have some model to work with, I will think of practical reasoning as proceeding as follows: it is a conscious mental process of tokening various propositions (with “assertoric force”?) in succession, with the rule that one can token a proposition in the succession if it’s a premise for that reasoning or if the preceding propositions in the reasoning are adequate reasons for it (either theoretical or practical; for present purposes, it doesn’t matter what account of reasons is adopted). After the succession of propositions, a decision or intention is issued, and we can call this the “conclusion” of that bit of practical reasoning. (Again, “drawing the conclusion” is governed by the rule that the succession must constitute an adequate reason for the conclusion.)

So can we, in this sense, just adopt a policy of taking a proposition as a premise in any practical reasoning where it is relevant? Here the answer would seem to be “yes.”

---

<sup>54</sup> This move, of course, fails: it would no doubt be irrational to take p as a premise in such a context, but Frankish’s datum is that it is *impossible* to decide to believe p when one has no evidence for p. The fact that it is irrational (which is near enough tautologous) hardly shows that it is impossible. But I’m not assuming it’s a datum that one cannot decide to believe what one has no evidence for, so this failing of Frankish’s account is immaterial for present purposes.

There is certainly no trouble in tokening a proposition in such a succession. One can token whatever conscious thoughts one wants.<sup>55</sup> Further, we can assume that forming an intention or making a decision is something one can do (“at will,” as it were).<sup>56</sup> So it would seem that I can, in fact, adopt a policy of tokening a proposition whenever it is relevant in a bit of reasoning and then draw the conclusion that reasoning supports (i.e. decide to act as the reasoning suggests one ought). To take an example, suppose I decide to use as a premise in future reasoning that I will receive funding for the upcoming academic year, something I am not at the moment particularly confident about. If I need to then decide whether it would be OK dip into my savings account to fund a small vacation, I can use the proposition that I will get funding next year to reason that I could afford the vacation. And on the basis of that reasoning, I can decide to dip into my savings.

No doubt, this model of practical reasoning is too crude to fully capture the phenomena. But if practical reasoning works anything like this, we have good reason to think that Frankish’s is right that we can adopt the relevant policy. Now let us consider theoretical reasoning. We can work with a model of theoretical reasoning that is similar to our model of practical reasoning: theoretical reasoning is a conscious mental process of

---

<sup>55</sup> Harman (1973) has famously argued, however, that real reasoning just is a process of changing belief. If Harman is right, Frankish cannot assume that one can decide to use a proposition as a premise in reasoning, because that would amount to a decision about how one will change one’s beliefs; this possibility is precisely what is at issue.

<sup>56</sup> The assumption is vexed, however: many contend that the formation of an intention or decision is not itself an action.

tokening various propositions (with “assertoric force”?) in succession using (syntactic?) rules from some deductive/inductive system. The final step in such a succession is the conclusion of that bit of reasoning.

Once again, it seems pretty clear that tokening a proposition in such a succession is something one can do. And since in the theoretical case, the conclusion is just the tokening of another proposition, that too is something one can unproblematically do. So, once again, it does seem that Frankish’s assumption is safe: in that sense, one can adopt a disposition to use a proposition in conscious reasoning by intending to do so from here on out, whenever that proposition is relevant.

But there is a complication in the case of theoretical reasoning that doesn’t arise for practical reasoning. Theoretical reasoning of the sort in question may or may not terminate in one’s believing the conclusion. I am currently unsure whether sweatshop labor in developing nations is, overall, good for the workers in those nations. I might decide, though, that from here on out I will use the proposition that it is in my reasoning (in an open-ended range of contexts). Suppose I’m now considering whether Apple, Inc. has an ethical duty to demand better working conditions in the Chinese factories where many of its products are made. Using the proposition that such jobs are ultimately good for the workers in China – which I was unsure about but decided to use in my reasoning – I reason my way to the conclusion that Apple has no such duty. Will that reasoning result

in my *believing* that Apple has no such duty? Probably not. The mere fact that I have decided to use some proposition in my reasoning doesn't mean I'm going to believe all the conclusions I can reach by reasoning from that proposition.

Or take a more clear example. I decide to start using in my reasoning the premise that there are thousands of tiny pink elephants on my desk (which is something, we can safely assume, I disbelieved before making that decision). Will I now believe the conclusions that I am able to draw from that premise – will I, for example, believe that my eyes are deceiving me when they tell me there are no such elephants, once I token the reasoning that terminates in that conclusion? Fortunately not.<sup>57</sup> Reasoning from premises I don't antecedently believe tends not to generate further beliefs.

What does this show? It shows, I think, something important about belief. To believe *p*, it is not sufficient to be disposed to use *p* as a premise in one's reasoning. It must also be the case that such reasoning tends to be *efficacious*: that is, it tends to result in one's believing the conclusions reached by such reasoning. Part of believing *p* is being disposed to believe things that one reasons follow from *p*. By showing that one can decide to use *p* in the reasoning, Frankish has not established that such reasoning will be efficacious. And without establishing *that*, it seems that he has not established that in deciding to use *p* as a premise, one has actually decided to *believe p*. And, it would seem,

---

<sup>57</sup> If I reason from this premise frequently enough, it is possible that I will start to believe it, and, hence, conclusions that follow from it.

Frankish cannot establish that one can adopt *that* disposition without assuming the very thing he's supposed to be showing: that one can adopt beliefs at will. For part of adopting that policy would be deciding to believe whatever one determine follows from one's premise.

But perhaps it is I who is begging the question, not Frankish. According to Frankish, belief *just is* being disposed to use p as a premise in one's reasoning. I have claimed that there is a further requirement: one must also be disposed to believe the conclusions of such reasoning. But on Frankish's conception of belief, this comes to the following: one must also be disposed to use the conclusions of such reasoning in any further reasoning. Can he establish that in deciding to use p in my reasoning, I am also deciding to use any conclusions I draw from p in any further reasoning?

On the conception of reasoning in consideration, it certainly does not follow from the fact that I have adopted the policy of using p in my reasoning, that I have adopted the policy of using any conclusions from such reasoning in further reasoning. I might adopt the former policy without adopting the latter policy. But Frankish, could, I suppose, complicate the policy that one must adopt in order to form a belief. One must decide that from here on out, one will not only use p whenever relevant, but will also use any conclusion q drawn from p whenever q is relevant. Insofar as we can adopt the policy to use p in our reasoning, it seems that we could adopt the more complicated policy of also

using conclusions drawn from  $p$ . So perhaps Frankish's argument survives my requirement that things you really believe are things that dispose you to believe further propositions that follow from them. (Though, I must confess, that it's not at all clear that we have enough control over our conscious reasoning to adopt, just like that, such an open-ended conditional disposition. Just by intending to, will I really dispose myself to use all of the crazy conclusions I can draw from my crazy premise?)

Some readers will find Frankish's account of belief incredible. Surely it's possible to adopt a policy of premising  $p$ , but to nonetheless not really believe  $p$ ! I am sympathetic, but if this is to amount to anything more than begging the question, we need to identify what more is required of belief. I now attempt to do just that.

### **Blocking the Gambit**

The dispositionalist gambit has not, to my knowledge, been discussed in the literature on doxastic voluntarism (though Frankish and Ginet are occasionally cited). Accordingly, I'd like to begin with two natural objections that I don't think are successful.

The first objection asks us to consider someone who, after reflecting on her evidence for  $p$ , finds herself without a belief either way about  $p$ . She then decides, as Frankish tells us is possible, to adopt the policy of using  $p$  as a premise in reasoning. The objection contends that, though she might *act as though* she believes that  $p$ , she doesn't

really *believe that* p. Frankish and Ginet confuse *pretending that* or *acting as if* you believe p with *actually* believing p. There is something seductive about this objection but, in its present form, it amounts to little more than begging the question. The central claim of Frankish and Ginet, though – and other dispositionalists about belief – is that believing *just is* “acting as if,” under some suitable specification of that notion. Belief is just a matter of having the right behavioral dispositions. If one’s behavioral dispositions vis-a-vis p are indiscernible from a believer of p, then one believes that p. One can, of course, reject a dispositionalist picture of belief.<sup>58</sup> But I want to grant dispositionalism. If we grant this, we must identify some specific disposition true believers possess that mere “Frankish-” (or “Ginet-”) believers lack.

An objection that fares better in this respect is inspired by L.J. Cohen’s conception of belief. According to Cohen, if one believes that p, then one is disposed to “feel it true that p,” when one considers the proposition that p (1989). As I understand Cohen, “feeling it true” is supposed to refer to a unique phenomenological property. Believing that p is being disposed to experience this property when one thinks about p. This is not plausibly a disposition one could voluntarily adopt – one doesn’t seem to have control over whether one’s thoughts have this property.

---

<sup>58</sup> Though one would have to argue additionally, I think, one can have the dispositions cited by Ginet and Frankish and yet still fail to meet some further condition on belief.



It seems to me, however, that this is an inessential feature of beliefs. Some beliefs have this feature, but there is reason to doubt that all do. It seems to me that many of the things we believe we believe unknowingly. Arguably, most people harbor racist beliefs in propositions that do not “feel true” to them. Less controversially, we sometimes reflect on our behavior and judge that the best explanation is that we believe *p*, though we had thought we didn’t believe *p*. That is to say that belief is not always introspectible, so it must not always have the phenomenology that Cohen ascribes it.<sup>59</sup> At any rate, the notion of “feeling it true” strikes me as too obscure to hang a theory of belief on; I would have thought that a major advantage of dispositionalism is that it could explain the apparent – but hard to describe – phenomenological properties of belief in terms of well-circumscribed behavioral dispositions.

In what remains, I will present a different line of objection – one that, it seems to me, establishes that the dispositionalist gambit fails and suggests that belief formation is, as a matter of psychological fact, always involuntary. To prepare the way, I’d like to begin with an illuminating quote from Ryle (emphasis added):

Certainly to believe that the ice is dangerously thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people’s assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition, and so forth. *But it is also to be prone to skate warily, to*

---

<sup>59</sup> Of course, one might suggest that in the case of unconscious beliefs, the disposition “to feel it true” is still present, but is masked by other factors. Thanks to Dan Howard-Snyder for making this point. This may be true of some unconscious beliefs, but I still see no good reason to suppose that for some things we believe, we are not at all disposed to feel it true. What’s more, zombies with beliefs seem to me perfectly possible.

*shudder, to dwell in imagination on possible disasters and to warn other skaters.* It is a propensity not only to make certain theoretical moves but also to make certain executive and imaginative moves, as well as *to have certain feelings.* (2009, pp. 117–118)

Ryle is here pointing to an important fact about beliefs: they guide not just our conscious, voluntary reasoning and behavior, but also our *involuntary* reactions (both physical and emotional) and our *unconscious* reasoning. Part of what it is to believe Smith a murderer is to be disposed, upon seeing Smith, to feel anger towards Smith, to have one's sympathetic nervous system trigger a stress response, or to get nervous if one sees Smith near one's daughter. Beliefs play an important role in the production and regulation of emotion and involuntary bodily reactions.

Perhaps even more importantly, beliefs play a role in automatic and unconscious reasoning. Consider an example from Audi (1994, 421). While deep in conversation with you, I may hear the siren of a fire-engine outside and unconsciously form the belief that a fire has broken out in town (later someone might say to me, "I haven't heard of any fires lately," and I might respond, "Well, I think there was one today.>"). To say that I form this belief unconsciously is to say that I do so without the proposition that there was a fire ever occurring as a conscious thought.

Now how did I form this belief? I must have unconsciously inferred that the sound I'd heard was the sound of a fire-engine, and from that inferred that there must be a fire in town. In making these inferences, I must have been appealing to background

beliefs of mine (e.g., that fire engines sound like this, that fire-engines sound their sirens when there's a fire). Indeed, a great many of our beliefs are formed unconsciously and result from unconscious inferences involving background beliefs.

Further, it seems that even in the case of conscious belief formation, one unconsciously relies on background beliefs. Take a tired example. Smith is considering hiring Green and reads one of Green's reference letters, which simply states, "Green is very punctual and his hand-writing is neat." One concludes the letter-writer doesn't think very highly of Green. The conscious reasoning may have been fairly simple:

1. All the letter-writer said is that Green is punctual and has neat hand-writing.
2. If the letter-writer thought highly of Green, he would have said more.
3. So the letter writer doesn't think very highly of Green.

But where did the second premise of this reasoning come from? Presumably, from a large number of background beliefs about letter-writers, conversational implicature, etc. One quickly, unconsciously, and automatically deployed such beliefs to spontaneously arrive at the second premise. Without those background beliefs being available for this sort of unconscious reasoning, one never would have arrived at the conclusion that Green's letter writer doesn't think highly of him. Reflection on aesthetic judgments or, say, judgments

about the mental states of others make the point very clear. We are typically unable, in such cases, to make explicit the unconscious reasoning we crucially rely on.

It seems to me that this suggests an important role beliefs play in our mental lives: they are available for use in unconscious reasoning. To put the point in dispositionalist terms, part of believing *p* is being disposed to use *p* as a premise in *unconscious* reasoning.

We can generalize from the above that believing that *p* requires: (i) being disposed to use it as a premise in unconscious reasoning; (ii) being disposed to experience certain spontaneous emotional responses and to have certain involuntary bodily reactions. This presents a real problem for the dispositionalist gambit. Frankish has given us no reason to think that adopting a policy of premising *p* will result in dispositions to experience all of the appropriate emotional responses and bodily reactions. Nor has he given us any reason to suppose that such a policy will result in a disposition to use *p* in unconscious reasoning. At the very least, then, if I am right about these further aspects of the dispositional profile of belief, then Frankish has failed to establish that we can decide to believe, even granting that he has established that we can decide to use *p* as a premise in reasoning.

Consider an example. Suppose I decide that from here on out, I will use in my conscious reasoning the proposition that fire-engine sirens mean nuclear war. If you and I

are involved in conversation and I hear fire-engines in the background, will I suddenly be filled with dread or unconsciously form the belief that nuclear war has broken out? Probably not. And this, I contend, is an important difference between believing that p and merely acting as if p.

In order to succeed, the dispositionalist needs that the dispositions involved in believing that p can be voluntarily adopted. We have seen two sorts of dispositions (from Ginet and Frankish) for which this is plausible. I claim, however, that neither of these dispositions is sufficient for believing. Neither entail being disposed to use p in unconscious reasoning or to undergo appropriate emotional and bodily reactions. If I'm right in claiming that these latter dispositions are required for belief, then extant implementations of the dispositionalist gambit fail. That is the primary conclusion of this paper.

But I think we have the materials for a stronger conclusion. The unconscious reasoning and emotional bodily responses that I am talking about are *involuntary* – that is uncontroversial. Now what this means, according to me, is that there is no action of unconsciously taking p as a premise, no action of breaking out in a sweat, no action of feeling anger towards someone, and so on. These are paradigm cases of things that

merely *happen* to one: one *finds oneself* unconsciously relying on p or breaking out in a sweat.<sup>60</sup> To get where I want to take us, I need something like the following postulate:

*Disposing Oneself Postulate:* If there is no intentional action-type  $\phi$ , then there is no intentional action-type of disposing oneself to  $\phi$ .

If the disposing oneself postulate (DOP) is true, then it follows straightaway that doxastic voluntarism is false. For it follows from this principle that there is no intentional action-type *disposing oneself to unconsciously premise p*; but so disposing oneself is part of forming the belief that p.

Is DOP true? It certainly seems plausible. And it meshes well with the examples with which we began thinking about voluntarism. Recall our British soldier who lost the use of his arms. For him, we said, there is no action of tying his shoes. Is there available to him an action of *disposing himself* to tie his shoes (when they're untied, say)? It doesn't seem so, precisely because he cannot tie his shoes. Further, DOP explains our intuitions about what sorts of dispositions are under our control. Recall that I cannot dispose myself to wake up at six in the morning. DOP explains why: I have available to me no action-type *waking up at six*. Waking up is something that happens to me, not

---

<sup>60</sup> Again, one can perform actions that result in these events, but those actions do not constitute a further intentional action of, say, breaking out in a sweat.

something I do.<sup>61</sup> On the other hand, I do have available to me the action-type *using p as a premise in conscious reasoning*. So, according to DOP, I can dispose myself to use p in conscious reasoning.

I propose, then, that we tentatively accept DOP. If we accept my claims about the role that belief plays in the regulation of our involuntary responses, then, we have good reason for thinking that doxastic voluntarism is false. That's not to say that this is anything like a necessary truth. It's a contingent fact that many of our emotional responses or the functioning of our sympathetic nervous system are involuntary. And it's a contingent fact that much of our reasoning is unconscious. Perhaps there are could be beings in which these aspects of their psychology are voluntary. The dispositionalist gambit shows that for those beings, doxastic voluntarism is true. But we're not like those beings; doxastic voluntarism is not true of *us*.

---

<sup>61</sup> This doesn't seem to be true of everybody. In the Ian Flemming novels, James Bond is able to dispose himself to wake up at specific times – he can decide that tomorrow he will get up at 5, no alarm clock necessary. Perhaps this is because James Bond can wake himself up: that is something he can do. That is what DOP postulates, at any rate.

### CHAPTER 3: ACCEPTANCE, ALIEF, AND IN-BETWEEN BELIEF

Sometimes our automatic and unconscious behavior suggests that we believe one thing, while our deliberative and conscious behavior suggests that we believe something else. Often, we're aware of this conflict. We experience it as an inclination to think not-p when we know that not-p is irrational, that p is what we should think. We try use p in our conscious reasoning, but we know that we can't control our unconscious responses, that they'll still reflect our inclination to think that not-p.<sup>62</sup> Let's call this phenomenon "double-mindedness." The question arises: what do we really believe when we're double-minded — p or not-p?

In a recent paper, Eric Schwitzgebel carves up the logical space as follows (2010):

*Pro-judgment view*: The subject believes that p and fails to believe that not-p (Gendler 2008a, 2008b; Zimmerman 2007).

*Anti-judgment view*: The subject fails to believe that p and instead believes that not-p (Hunter 2011).

---

<sup>62</sup> I'm deliberately using the fudge word "think" here because I want to remain neutral for the moment on the correct ascription of propositional attitudes.



*Shifting view*: The subject shifts between believing p and believing that not-p (Elga 2005; Rowbottom 2007).

*Contradictory view*: The subject believes both p and not-p (Gertler forthcoming; Sommers 2009).

All of these views assume that there is a correct and determinate way to ascribe beliefs in these cases (Hunter isn't actually committed to this – he's happy to allow that some of these may be cases of in-between belief). Schwitzgebel rejects this assumption and offers the following view (2010):

*In-between-belief view*: The subject sort of believes p and sort of believes not-p (in the way that *Seinfeld's* George Costanza is sort of bald and sort of not-bald).

While useful, this schema has the misleading implication that settling what a subject believes is all there is to understanding representational conflict. But that is to neglect a more general question: which mental states explain our behavior in cases of representational conflict? When we ask this question, we leave open the possibility that something beyond belief is needed to fully understand representational conflict.

This paper proceeds as follows. I will illustrate the phenomenon of representational conflict with some common sense examples inspired by the work of Keith Lehrer. These examples share certain features that will define our explanatory

task. I will then argue that we need more than belief to explain the examples. Next I argue that Gendler's notion of alief is not of much help here. This motivates my introduction of a new propositional attitude called "acceptance." Some recent work in psychology demonstrates the empirical validity of both the phenomenon of representational conflict and the attitude of acceptance. I conclude with some reflection on what this means for epistemology.

### **Some Examples**

Much of the existing literature has focused on cases like Schwitzgebel's *Juliet the implicit racist* (2010):

Many Caucasians in academia profess that all races are of equal intelligence. Juliet, let's suppose, is one such person, a Caucasian-American philosophy professor. She has, perhaps, studied the matter more than most: She has critically examined the literature on racial differences in intelligence, and she finds the case for racial equality compelling. ... And yet Juliet is systematically racist in most of her spontaneous reactions, her unguarded behavior, and her judgments about particular cases. When she gazes out on class the first day of each term, she can't help but think that some students look brighter than others - and to her, the black students never look bright. When a black student makes an insightful comment or submits an excellent essay, she feels more surprise than she would were a white or Asian student to do so, even though her black students make insightful comments and submit excellent essays at the same rate as do the others. This bias affects her grading and the way she guides class discussion. She is similarly biased against black non-students. When Juliet is on the hiring committee for a new office manager, it won't seem to her that the black applicants are the most intellectually capable, even if they are; or if she does become convinced of the intelligence of a black applicant, it will have taken more evidence than if the applicant had been white. When she converses with a custodian or cashier, she expects less wit if the person is black. And so

on. Juliet could even be perfectly aware of these facts about herself; she could aspire to reform; self-deception could be largely absent. We can imagine that sometimes Juliet deliberately strives to overcome her bias in particular cases. She sometimes tries to interpret black students' comments especially generously. ...

Juliet's behavior is puzzling. What is her attitude to the proposition that the races are equal? What is her attitude to the proposition that blacks are less intelligent than whites? Correct answers to these questions, we hope, will explain Juliet's conflicted behavior. This sort of case, however, is not exactly what I have in mind when I talk of representational conflict. Notice, for example, that when Juliet considers the proposition that blacks are less intelligent than whites, she's not at all tempted to endorse it. Her spontaneous reaction is that the proposition is demonstrably false (not to mention morally repugnant). This is not a case of finding oneself with a strong inclination to think that  $p$ , while knowing that  $p$  is irrational.

At the end of the quoted passage, Schwitzgebel says that "we can imagine that sometimes Juliet deliberately strives to overcome her bias in particular cases." This insight starts to get us into the territory I want to consider. Imagine Juliet is considering the academic performance of a particular black student she has always regarded as less bright than the rest of the class. On careful review of the evidence, it becomes clear that this student is actually quite gifted, perhaps the best in the class. It's not too hard to imagine that Juliet still finds herself inclined to think that this student is

average at best, even though she knows this is irrational. What happens next? Let's consider some examples.

*Ian and Mueller-Lyer drawings*

I am familiar with the Mueller-Lyer illusion. I've seen dozens of very impressive illustrations of the effect. I know that, if drawn well, the lines *really look* like they're different lengths, even though they're not. The funny thing is, when I see a new and particularly well-drawn example, my first reaction is to think "This one can't be real; it's *too* good. These lines must be different lengths." This is accompanied by a strong desire to measure the lines to confirm my judgment. I'm usually quick to admonish myself – this drawing is plainly no better than countless others I've seen that have all been the real deal; it's from a reputed psychology text; etc. If someone asks me, I'll tell them that the lines are the same length. I'll bring it to class to illustrate a point to my students. Etc. But when no one's looking, I might give in to temptation and measure the lines. Were you to offer me a modest bet, I'd hesitate before forcing myself to bet that the lines were the same length.

*Dana and the skeletons in her closet*

When Dana was a child, she believed that there were dangerous skeletons in her closet that presented a real night-time threat. Of course, now, as a young adult, Dana is well

aware of the reasons for thinking that there are no dangerous skeletons in her closet – how could they be animate without muscles and a nervous system? She'll happily tell you that there are no dangerous skeletons in her closet. But sometimes she wakes up in the middle of the night and is cold. She needs to get the extra blanket out of the closet and visions of menacing skeletons occupy her attention. It often takes her several minutes of rehearsing the evidence against skeletons to get up and get the blanket. Sometimes, when she only half wakes up, Dana doesn't get this far and falls back asleep deciding to deal with the cold rather than skeletons.

*McCoy the drunk driver*

McCoy is well aware of the evidence that drunk driving is, statistically, extremely dangerous. He's also aware that he has no reason to regard himself as the exception to the statistics (he's also aware that any tendency to regard himself as an exception is almost certainly due to a common and well-understood psychological bias), and in fact has good reason to think that he himself drives dangerously when drunk. When going out drinking, he takes a cab or finds a designated driver in advance. If one of his past episodes of drunk driving comes up, he's quick to express regret at such dangerous behavior. And yet, after a few drinks, McCoy usually thinks that he could drive perfectly safely at the moment. And even when not drunk, McCoy usually finds himself tempted to think that he's actually a capable drunk driver, that drink doesn't

affect his motor skills and reaction times like it does others. He has to consciously disavow these thoughts and deciding to take a cab or finding a drunk driver, though he does it, requires real willpower.

*Green the racist juror*

Green harbors racist beliefs as a result of his intolerant upbringing. He has since come to recognize that these beliefs are both unfounded and pernicious. But after a lifetime of reinforcement by racist peers, they are too deeply entrenched to be changed. Green is a juror in a murder trial. A black man has been accused of murdering a white woman. Green immediately and unconsciously believes that the black man is guilty. But he recognizes this belief for what it is and patiently watches the trial. The defense has a knock-down argument that the defendant is innocent. The prosecution only offers slander of the defendant's character. Green recognizes that the evidence conclusively establishes the defendant's innocence. He admirably votes to acquit. Whenever the question comes up, Green asserts that the defendant was innocent. This all requires great effort on Green's part – he still finds himself inclined to think that the defendant is guilty. After the trial, whenever Green hears the acquitted defendant's name or sees a picture, he wells up with anger (though he's typically quick to consciously remind himself that such anger is misdirected). When drinking heavily and in bad company,

Green might even go so far as to call the defendant a murderer (something he'll very guilty about the next morning).<sup>63</sup>

These are paradigm cases of representational conflict. These cases are, I hope, familiar (perhaps *too* familiar). But familiar though they are, they are also quite puzzling. They require philosophical explanation. In what follows, I will focus on Green the racist juror, but what I say generalizes to all these cases.<sup>64</sup>

### **The Explanatory Challenge**

We want to attribute mental states to Green that explain his behavior. Let's identify the features of Green's behavior that we want explained. We can distinguish between Green's *implicit* and *explicit* behavior. His implicit behavior includes breaking out in a sweat when he sees the defendant and his welling up with anger when he hears the defendant's name. This behavior *implicitly* suggests an attitude about the defendant's guilt. His explicit behavior includes his vote to acquit, his sober assertions that the defendant is innocent, etc. It *explicitly* suggests an attitude about the defendant's guilt. Further, let's distinguish between *guilty-behavior* (suggestive of an attitude that the

---

63 Some readers might notice that this case is similar to one of Lackey's (2007). Lackey uses the case as a counterexample to the thesis that knowledge is the norm of assertion. Her argument assumes that knowledge entails belief. If I'm right, then though Green does not believe that the defendant is innocent, he nevertheless knows it. This undermines Lackey's counterexample.

64 I'm not, though, committed to a particular view about Juliet and her attitude towards the proposition that the races are equal. That case (and many others found in Gendler) could very well require different treatment.

defendant is guilty) and *innocent-behavior* (suggestive of an attitude that the defendant is innocent). Then the following are true of Green:

1. Green's implicit behavior is guilty-behavior.
2. Green's explicit behavior is innocent-behavior when he is being careful and reflective, but guilty-behavior when not (like when he's drunk).
3. Green's innocent-behavior is *effortful* – he finds guilty-behavior natural and must consciously suppress it (though this likely becomes easier with time).

The mental states we attribute to Green should explain these features of his behavior. Simply attributing a belief that the defendant is innocent leaves feature 1 unexplained. If Green really believes the defendant is innocent, why is a large class of his behavior suggestive of a belief that the defendant is guilty? If, on the other hand, we simply attribute to Green the belief that the defendant is guilty, then we are unable to explain features 2 and 3. If Green really believes that the defendant is guilty, then why on earth does he vote to acquit? Why does he tell everyone that Green was falsely accused?

Green's behavior is conflicted in a way that suggests mental conflict.

Attributing mental states that explain this conflict is the explanatory challenge raised by cases of representational conflict.

*Belief-Only Explanations*



Recall that two of the options on the table are the pro-judgment view and the anti-judgment view. If we think of these as offering complete explanations of the cases, then they are non-starters for the reasons we have just seen.<sup>65</sup> If belief is all we have recourse to, this leaves the shifting, contradictory, and in-between belief views. The appeal of these views is their economy: it would be nice to explain representational conflict without inflating our mental ontology. Unfortunately, such economy leaves us explanatorily impoverished; or so I will argue in this section.

Suppose we say that Green sometimes believes that the defendant is guilty and other times believes that he's innocent. This is the shifting belief view.<sup>66</sup> The trouble with this view is that it doesn't explain why Green's implicit behavior is *invariably* guilty-behavior. For at those times, Green is currently believing that the defendant is innocent (and failing to believe that he is guilty), he is still disposed to display implicit guilty-behavior. This becomes particularly pressing when we notice that this implicit guilty-behavior can occur at the same time Green is displaying explicit innocent-behavior. (Someone points at the defendant and asks Green "So he's innocent?" As Green looks at the defendant, he feels a flash of anger and breaks out in a light sweat

---

65 This is why Gendler supplements the notion of belief with alief – aliefs are supposed to explain Green's implicit behavior. Hunter isn't directly concerned with cases where one's overall pattern of behavior is inconsistent. Zimmerman, however, cannot explain the phenomena.

66 Hume seemed to hold this view of himself with respect to his skeptical philosophical beliefs. Did he believe that there were physical objects? He did not, he said, when he was in his study, but he did when he returned to his friends in the billiard room. Elga (2005) attempts to extend this sort of explanation to cases of representational conflict similar to mine.

as he determinedly says, “Yes, he’s innocent.”) This is inexplicable on the shifting belief view: when Green is currently believing that the defendant is innocent, we should not expect behavior suggestive of belief that the defendant is guilty; that, however, is exactly what we find. The flipside, of course, is that when Green is currently believing that the defendant is guilty, we will still find that his deliberative explicit behavior is innocent-behavior. This, too, is inexplicable on the shifting belief view.

The contradictory belief view is equally unsatisfying. On this view we attribute to Green both the belief that the defendant is guilty *and* the belief that the defendant is innocent (and at the same time). The trouble with this view is that the two beliefs would have to have inexplicably different functional roles. Green’s belief that the defendant is innocent has no effect on Green’s implicit behavior, but it trumps his belief that the defendant is guilty when it comes to (deliberative) explicit behavior. His belief that the defendant is guilty, on the other hand, guides his implicit behavior and his (non-deliberative) explicit behavior. Further, activation of his guilty-belief is more or less automatic, while activation of his innocent-belief requires effort. If both states are beliefs, then it is mysterious why they should have such radically different effects on behavior and why one should be automatic and the other effortful.

*In-Between Belief*

A more sophisticated explanation invokes Schwitzgebel's notion of in-between belief. Suppose we think of belief in dispositional terms – to believe *p* is to have certain behavioral dispositions (where “behavior” is understood broadly enough to include mental activity). Of course, there will be different accounts of just what sorts of dispositions are required. But on any such account, it will be possible to have someone with most, but not all, of the relevant dispositions. Of such people, it will be tempting to say that they believe *p* and it will be tempting to say they disbelieve *p*. But neither attribution will be determinately correct. The best we can do is to say such people are in-between belief and disbelief – they sort of believe *p* and they sort of disbelieve *p* (Schwitzgebel 2010: 8).

Schwitzgebel has deployed this notion of in-between belief to explain the case of Juliet the implicit racist. On the question whether Juliet really believes the races are equal, the best we can do is say she only sort of believes this – she also sort of believes that whites are intellectually superior. She is disposed to (sincerely) assert that the races are equal, to marshal evidence supporting this, to support relevant political causes, etc. However, she is not disposed to treat the races equally – she is disposed to treat whites as smarter than blacks.

Perhaps Schwitzgebel would be tempted to offer the same explanation of Green the racist juror. As the case is described, Schwitzgebel may say has many of the

dispositions characteristic of believing that the defendant is guilty, but lacks too many to count as fully believing this. There is, for example, his disposition to get angry when he thinks about the defendant, his disposition to intuitively find his guilt plausible, his disposition to assert that the defendant is guilty when he's blind drunk, etc – all of that is how you behave when you believe the defendant is guilty. On the one hand, Green didn't vote "guilty," when he's carefully reasoning about the evidence he doesn't think it supports the defendant's guilty, he does not ordinarily assert that the defendant is guilty, and so on. And for these same reasons, one might think that Green has many of the dispositions characteristic of believing that the defendant is innocent, but lacks too many to count as fully believing this. On this line of thought, then, we cannot accurately say that Green believes the defendant is innocent and we cannot accurately say that he believes the defendant is guilty. Green is in-between belief.

I have two objections to this way of describing Green's mental states: (i) I think Green *does* have all the dispositions characteristic of believing the defendant is guilty (that's why his innocent-behavior is effortful); (ii) there's a systematic cleave between the implicit and the explicit in Green's behavior, and labeling him as an in-between believer doesn't really explain this.

To conclude that Green lacks some dispositions characteristic of believing that the defendant is innocent is to move too quickly from what Green actually does to what

Green is disposed to do. From the fact that he votes to acquit, it doesn't follow that Green wasn't at all *disposed* to vote to convict. An important feature of dispositions is that they can be *masked*: a glass packed in styrofoam will not break when dropped even though it is fragile. As I understand the case, Green *is* disposed to vote to convict. It's just that this disposition was masked by his belief that the evidence compellingly supports the defendant's innocence, and his decision, on the basis of the evidence, to take the defendant's innocence as a premise in reasoning. But the fact that it's difficult for him to vote to acquit – he's tempted to vote guilty – suggests that he is in fact disposed to vote that the defendant is guilty. Green, it seems to me, has all of the dispositions characteristic of believing that the defendant is guilty, but they are masked by his policy decision to reason as though the defendant is innocent. These dispositions, however, are allowed to shine through when Green is drunk or careless. So I think we can quite accurately say that Green fully believes that the defendant is guilty.

Green's innocent state is a bit different. It does seem true that Green has many dispositions characteristic of a belief that the defendant is innocent, but also lacks many of the dispositions characteristic of such a belief. He is disposed to say that the defendant is innocent, but only when he is sober. People who really believe the defendant is innocent are disposed to assert as much even when they are drunk. Further, he does not seem disposed to implicitly behave in the ways someone who

really believed the defendant is innocent would. Green's innocent state might really qualify as an in-between belief in Schwitzgebel's sense. The trouble with this, according to Schwitzgebel, is that it isn't possible if Green fully believes that the defendant is guilty, as I've suggested he does. For that would suggest possessing contradictory dispositions, which Schwitzgebel takes to be impossible (and for good reason).

If we think about Green's dispositions carefully, though, we will see that they are not contradictory. Consider the following pair:

1. Green is disposed to assert that the defendant is guilty when he's being careful and reflective *unless* he decides to override this disposition.
2. Green is disposed (when being careful and reflective) to override his inclination to call the defendant guilty and assert instead that the defendant is innocent.

Many dispositions associated with belief are such that we can voluntarily override them. Green's innocent state disposes him to override all such dispositions. However, there are other dispositions associated with belief that we cannot voluntarily override (like Green's tendency to break out in a light sweat when he sees the defendant).

Green's innocent state does not dispose him to override such dispositions.

So it may be that Schwitzgebel thinks of Green as fully believing that the defendant is guilty while sort of believing that he is innocent. But this won't, I want to stress, explain a very interesting (and systematic) difference between Green's innocent state and his guilty state. Absent conscious effort, Green's guilty state will guide all behavior. His innocent state only comes into play when Green is being careful and this requires conscious effort. Why this difference? Why doesn't Green's innocent state sometimes automatically guide behavior, perhaps with the guilty state – through conscious effort – overriding? These are questions that we can't readily answer if we think of Green's innocent state as a mere "almost-belief."

Perhaps this last point is less than decisive. As we'll see momentarily, however, Green's innocent-state really isn't a belief at all, borderline or otherwise. I'll argue for this claim in the next section. First, however, I'd like to summarize the discussion thus far.

### *Summary of Section*

In brief, the problem with belief-only views is this. Green's state with respect to the proposition that the defendant is guilty (call it his *guilty-state*) plays a different functional role from his state with respect to the proposition that the defendant is innocent (his *innocent-state*). So we seem to be dealing with two different types of mental states – Green's innocent-state is just *a different sort of thing* from his guilty-

state. The innocent-state is partially characterized by the fact that it guides implicit behavior and explicit behavior; the guilty-state is partially defined by the fact that it guides reflective explicit behavior and can, with effort, override a conflicting implicit state in this regard.

We have seen that there is good reason to view Green's guilty state as a belief – it seems to have the right dispositional profile. Green's innocent state, on the other hand, it would be better to call it a paradigm *something else*. We should, then, look for a new type mental state that has the right functional role to serve as Green's innocent state. What we will find is that this new type of mental state, while similar to belief, is a propositional attitude in its own right. I call the attitude *acceptance* – Green believes that the defendant is guilty but accepts that he is innocent.

Before explaining my own view, however, we will consider an account that views Green's *innocent-state* as a belief and his guilty-state as something else: an alief.

### **Alief**

In a recent series of influential papers (2008a and 2008b), Gendler argues that that the “classic” cognitive and conative attitudes cannot adequately explain large swaths of behavior. Why do you hesitate to put a piece of feces-shaped chocolate in your mouth, jump when a monster appears on a movie screen, or freeze with fear on the glass



walkway extending over the Grand Canyon? After all, you believe that the chocolate is just chocolate, that the monster is fictional, and that standing on the walkway is perfectly safe. Your beliefs and desires don't explain this behavior. Gendler introduces a new category of mental state, *alief*, that is supposed to be up to the task. Imprecisely (2008b: 557):

To have an alief is, to a reasonable approximation, to have an innate or habitual propensity to respond to an apparent stimulus in a particular way. More precisely (2008a: 642):

A paradigmatic alief is a mental state with associatively linked content that is representational, affective and behavioral, and that is activated – consciously or nonconsciously – by features of the subject's internal or ambient environment. Aliefs may be either occurrent or dispositional.

Seeing the feces-shaped chocolate triggers an alief with the contents <Feces (representational), Gross! (affective), Don't eat! (behavioral)>; seeing the monster triggers an alief with the contents <Monster, Yikes!, Activate fight or flight!>; seeing the Canyon floor thousands of feet beneath you triggers an alief with the contents <High, Yikes!, Don't move!>. The contents of these aliefs are not linked by beliefs or inferential reasoning. The link is due to a lifetime of associations.

According to Gendler, aliefs are *\_a\_ssociative*, *\_a\_utomatic*, *\_a\_rational*, *\_a\_ffect-laden*, and *\_a\_ction-generating* (2008b: 557). Alief is not supposed to be a propositional attitude: they have no propositional content. For this reason, aliefs are

not available to (or affected by) inferential reasoning.<sup>67</sup> This, I take it, is why Gendler says that alief is arational. No amount of reflection on the impressive engineering of the Grand Canyon walkway will eliminate the alief that has you frozen with fear. There's no *rational* process that will change an alief – they can only be changed by developing new associations. Incidentally, Gendler regards belief as *constitutively* affected by rational processes; belief *just is* the result reasoning about evidence.

I agree with Gendler that the notion of alief is helpful in explaining many puzzling facets of human behavior. But the question here is whether alief can explain paradigm cases of representational conflict. If we add alief to our quiver, are we able to explain Green's behavior? I think we are not. Let us first consider what mental states Gendler might attribute to Green. The state that influences his innocent behavior is, however, the result of reflection on the evidence; according to Gendler, that makes it a belief. Presumably, then, Gendler would be tempted to say:

1. Green's innocent state is a belief.

In other words, Green believes that the defendant is innocent. Unlike his innocent state, however, Green's guilty state is unaffected by reflection on evidence. Gendler cannot, then, regard it as a belief. Perhaps we're looking at an alief. Of course, no single alief can explain the wide variety of Green's guilty-behavior. We would need to

---

<sup>67</sup> I'm using "reasoning" in a restricted form here: the sort of reasoning in question involves transitions between contents governed by syntactic or logical relations.

attribute to Green many aliefs: perhaps an alief with the content <Black man, Disgust, Make a face>; an alief with the content <Black man with white woman, Grr!, Get him! >; and so on. At any rate, the suggestion is that:

2. Green's guilty states are aliefs.

1 and 2 taken together offer a potential explanation of Green's behavior. There are two main problems with this suggestion: (i) it is implausible to attribute to Green a *belief* that the defendant is innocent, and (ii) Green's guilty state seems to play an inferential role in Green's thinking – it seems to have propositional content.

First problem first. Green, I said, arrived at his innocent state through an exercise of direct voluntary control. Upon considering the evidence, Green judges that it establishes the defendant's innocence. True enough, this judgment was not voluntary in any sense – it resulted from Green's voluntary decision to think about the evidence, but once the process of thinking about the evidence was underway, Green couldn't help but judge that the evidence supported the defendant's innocence. The thing about occurrent judgments, however, is that they don't guarantee that the agent will form any corresponding dispositional state (a belief, say). It's all well and good to go through an argument and token the conclusion with some sort of assertoric force, but to think that such an exercise automatically will have an impact on one's behavioral dispositions

would be unduly naive.<sup>68</sup> Green knows this. He knows that, though he's now attending to the evidence and judging that the defendant is innocent, if he doesn't *do something*, he'll go right back to acting and reasoning as though the defendant is guilty. What does he do? He forms the *resolution* to take the defendant's innocence as a premise in his (conscious) theoretical and practical reasoning. If Green is continent – able to follow through on his resolutions – he will now be disposed to take the defendant's innocence as a premise in his reasoning. This will undergird dispositions to assert that the defendant is innocent in relevant contexts, to vote for his acquittal, etc. This dispositional state is Green's innocent state. Insofar as we can control our conscious reasoning, then, Green's innocent state is under his direct voluntary control. Belief, of course, is not under our voluntary control in this way, as Gendler herself emphasizes (2008a: 651). So Green's innocent state is not a belief state (though it *may* result from a belief about what the evidence supports).

The second problem is that Green's guilty state seems to have propositional content – the content that the defendant is guilty. Green is disposed to reason with the proposition that the defendant is guilty. He might find himself thinking, “Tina wants to go to the dance tonight, but I shouldn't let her go out with that rapist on the streets.” He may correct that reasoning almost instantaneously, or manage to suppress it entirely, but he still finds it tempting. That's strongly indicative of a state with a

---

<sup>68</sup> For some interesting discussion of the distinction between belief and occurrent judgment, see Lycan (2000).

propositional content, which means it's not an alief. And notice how this is different from, say, the feces-shaped chocolate case. You're not at all disposed to reason with the proposition that this is a piece of feces. Whatever Green's guilty state is, it's available for use in reasoning and seems to have propositional content. That state can't be an alief, and that state, it seems to me, is crucial in explaining Green's behavior. Some of Green's guilty behavior might be explicable in terms of aliefs, but that can't be the whole story.

Can Gendler just concede the latter point and say that Green believes that the defendant is guilty? That seems like the correct thing to say to me, but Gendler can't accept it. Green's guilty state is impervious to reflection on the evidence – try as he might, he just can't get rid of it. According to Gendler, this disqualifies it from being a belief. But even if Gendler could agree that Green believes that the defendant is guilty, we would still lack an account of Green's innocent state (it's voluntary and so not a belief, and even if we thought some beliefs could be voluntary, I've already argued that ascribing conflicting beliefs to Green doesn't explain the data).

Alief may be a useful explanatory tool for some aspects of human behavior, but it's not what we need to understand a paradigm case representational conflict like Green the racist juror.

### **Acceptance**

I've argued that Green's guilty-state is a belief and that his innocent state is not.

Green's innocent state is a propositional attitude that isn't explicable in terms of belief.

For reasons that will become clear, I will call this attitude *acceptance*. What is

acceptance? We can characterize it roughly as follows:

1. It is under direct voluntary control.
2. It guides deliberative explicit behavior, overriding conflicting beliefs if necessary.

But we can identify acceptance more directly. To accept that *p* is to have adopted a policy of taking *p* as a premise in one's theoretical and practical reasoning. We have control over what we premise in our (conscious) reasoning, and so we can decide to adopt such a policy. That explains why acceptance is under voluntary control and why it guides explicit behavior. It also explains why this explicit behavior is effortful – it takes an act of will to premise something we don't believe. It also explains why acceptance does not affect implicit behavior – such behavior is not within the purview of our policy decisions.

To be in the state of accepting that *p* is to be disposed to take *p* as a premise in explicit theoretical and practical reasoning as the result of a policy decision. Does deciding to adopt such a policy guarantee that one is in the dispositional state of acceptance? No. Someone may decide to quit smoking without thereby becoming a non-smoker. Policy decisions can fail due to weakness of will. Still, when all goes well,

deciding to adopt the policy of reasoning as though p will result in one's actually being disposed to reason as though p.

Supposing that p also involves a policy of reasoning as though p. Acceptance is similar to supposition. However, when one supposes p, it is always limited to a context: one supposes during a debate, or in one's research, or as a prosecutor. Acceptance, though, spans across context. To accept that p is to be disposed to reason as though p, *whatever the context*. In this respect, acceptance is like belief.

Of course, one can decide to adopt a policy of reasoning as though p for all sorts of reasons. One might think it prudent or morally obligatory, for example. It is a hallmark of Keith Lehrer's epistemology, however, that one accepts that p for a distinctively epistemic purpose: for the sake of accepting what is true and not accepting what is false (2000a; 2000b; 2003). According to Lehrer, part of the functional role of acceptance is that it is a state arrived at via rational reflection on epistemic reasons.

Here I take a more ecumenical approach. On my usage, acceptance is a state one can adopt for a variety of reasons, not all of them epistemic.<sup>69</sup> That leaves open room for an ethics of acceptance: in general, what ought we to accept?<sup>70</sup> But that question is one

---

<sup>69</sup> This raises a general point. A variety of propositional attitudes very similar to the present notion of acceptance have been introduced by philosophers. A taxonomy and evaluation of these accounts would be helpful, but beyond the scope of this paper. I am seeking only to defend the claim that we need something like the present notion of acceptance to understand representational conflict.

<sup>70</sup> I suspect that we have a prima facie ethical duty to accept p only if it would be epistemically rational for us to believe that p, but that is not a trivial result on the present account of acceptance.

for another day. Suffice it to say that I am not here imposing limits on how one gets into a state of acceptance.

Still, what concerns us at the moment is accepting *p* for epistemic reasons. One does not ordinarily accept what one has epistemic reason to accept, for one generally *believes* what one has epistemic reason to accept. If one is already disposed to reason as though *p*, one does not need to adopt a policy to that effect. However, sometimes we find ourselves disbelieving *p* *in spite of* thinking that the evidence for *p* is conclusive. That is, we sometimes find ourselves with irrational beliefs.<sup>71</sup> It is in these cases that one might decide to accept that *p* on the basis of the evidence.

This, I suggest, is precisely how we should think of Green's innocent-state. Even after reflecting on the evidence, Green found himself thinking that the defendant was guilty. He knew that this was irrational, but he couldn't help it. So Green made a policy decision: he decided not to let this belief affect his behavior. He would reason and act as though the defendant was innocent, in accord with the evidence. Green accepted that the defendant was innocent even though he couldn't shake the belief that he was guilty.

This attribution of mental states meets the explanatory challenge that we set at the beginning of this paper. The puzzling features of Green's behavior were:

---

<sup>71</sup> Of course, it is controversial that we can simultaneously believe *p* and judge that our belief is irrational. For defense of the possibility, see Hunter (2011), Mele (1987: chapter 8), and Evans (unpublished).



1. Green's implicit behavior is guilty-behavior.
2. Green's explicit behavior is innocent-behavior when he is being careful and reflective, but guilty-behavior when not (like when he's drunk).
3. Green's innocent-behavior is *effortful* – he finds guilty-behavior natural and must consciously suppress it (though this likely becomes easier with time).

Since belief guides implicit behavior, and Green believes that the defendant is guilty, it is clear why 1 is true. Of course, belief also guides explicit behavior and this explains 3. But since acceptance trumps belief when it comes to reflective explicit behavior, and Green accepts that the defendant is innocent, we see why 2 is true of Green.

At this point, one might worry that my description of Green simply stipulates what I'm attempting to argue. I hope that my description of Green was folk-psychologically plausible enough to allay this suspicion. But we needn't settle for plausible folk psychology. There is a program in empirical psychology that has been contending with just the sorts of phenomena I'm discussing and that has proposed an explanatory model strikingly similar to mine.

### *Motivated Overriding*

The program I have in mind draws a distinction between *implicit* and *explicit* mental processes and states. What's interesting for our purposes is that "[r]esearch on such

disparate topics as perception, motor learning, personality, attitudes, and self-esteem reveals a frequent discordance between implicit and explicit measures of internal states” (Wilson & Dunn 2004: 504).

Wilson & Dunn define implicit states as those that (2004: 104):

1. have unknown origins;
2. are activated automatically;
3. influence implicit responses, namely, uncontrollable responses and ones that people do not view as an expression of their attitude and thus do not attempt to control.

Explicit states, on the other hand, are typically activated consciously and deliberately and influence explicit responses (verbal reports of one’s states, for example, deliberative action, for another). Explicit states are also typically the result of conscious reflection, and so their origins are accessible.

To give a concrete example, a person might have conflicting implicit and explicit motivations. She might verbally (and sincerely) deny that acquisition of power motivates her, but her implicit behavior might strongly suggest that power does motivate her. We might think of such a person as self-deceived. But, alternatively, we might think the

situation more complicated: when engaging in careful practical deliberation, power might not motivate her; when behaving less deliberately, however, it will.

Seeking to understand discordance between implicit and explicit reactive attitudes (as measured by conflicting performances on implicit and explicit attitude tests), Wilson & Dunn have proposed what they call the *Dual-Attitude* model (2004). This model predicts that (104):

1. Explicit attitudes (EA) and implicit attitudes towards the same object can coexist in memory.
2. When dual attitudes exist, the implicit attitude (IA) is activated automatically, whereas the explicit one requires more capacity and motivation to retrieve from memory. When people have the capacity to retrieve EA, it can override IA, such that they report EA. When people do not have the capacity and motivation to retrieve EA, they report IA.
3. Even when the explicit attitude has been retrieved from memory, IA influences implicit responses, namely, uncontrollable responses (e.g. some nonverbal behaviors) or responses that people do not view as an expression of their attitude and thus do not attempt to control.

4. Explicit attitudes change relatively easily, whereas implicit attitudes, like old habits, change more slowly. Attitude-change techniques often change explicit but not implicit attitudes.
5. Dual attitudes are distinct from ambivalence and attitudes with discrepant affective and cognitive components. Rather than experiencing a subjective state of conflict, people with dual attitudes report the attitude that is most accessible.

There are two variables that can generate interestingly different kinds of dual-attitudes:

(i) whether the agent has access to the implicit attitude, and (ii) whether cognitive capacity and motivation is required to override the implicit attitude with an explicit one (Wilson & Dunn: 105):

#### Varieties of Dual Attitude

	<b>Awareness of IA?</b>	<b>Capacity/Motivation req. to Override?</b>
Repression	No	Yes
Independent Systems	No	No
Motivated Overriding	Yes	Yes
Automatic Overriding	Sometimes	No

Repression and motivated overriding are of chief interest for us. Repression occurs when the implicit attitude is blocked from conscious access because it is anxiety-inducing

(Wilson & Dunn: 106). The existence of repression is controversial and the existing data is inconclusive. Motivated overriding, on the other hand, is fairly well empirically established. In cases of motivated overriding, the agent is aware of the implicit attitude, but regards it as illegitimate and unwanted. When not under heavy cognitive load, the agent can override the implicit attitude in controlling her explicit behavior (though implicit behavior will still be influenced by the implicit attitude). When no overriding occurs, implicit attitudes guide explicit behavior.

Green the racist juror is a paradigm case of motivated overriding. He is aware of an implicit attitude that the defendant is guilty and capacity and motivation are required to override this implicit attitude with his explicit judgment that the defendant is innocent. He regards his implicit attitude as illegitimate and unwanted. The implicit attitude guides implicit behavior even when explicit behavior is being guided by the explicit attitude. What sorts of states are these implicit and explicit attitudes? I have argued that the implicit attitude is a belief and the explicit one is an acceptance.

Belief fits the profile of an implicit state pretty well. It is “activated” automatically; it guides implicit behavior; its origins are often unknown.<sup>72</sup> Acceptance, on the other hand, has all the hallmarks of an explicit state: it is subject to direct voluntary control and accessible to consciousness (I’m inclined to think we are aware of

---

<sup>72</sup> Of course, we often think we know why we believe what we do, but such judgments are highly fallible.

its origins, but that might just be wishful thinking on my part). I think, then, that both my description of the case and my psychological explanation are empirically plausible.

I have articulated a propositional attitude that heretofore has not received adequate attention: acceptance. Acceptance is distinct from belief. One can accept what one does not believe and one can believe what one does not accept. Accepting that *p* is being disposed to take *p* as a premise in (conscious) theoretical and practical reasoning. Since one can decide to be so disposed, one can decide what to accept. One can decide to accept *p* because the evidence supports *p*. This is of chief interest when one (a) believes that the evidence supports *p* but (b) due to irrationality, one cannot get oneself to believe that *p*. In such situations, one can at least accept that *p*. By doing this, one will at least bring one's explicit behavior into line with the evidence.

I have argued that this is what is going on in cases of representational conflict. An otherwise puzzling pattern of behavior is made intelligible by recognizing the difference between belief and acceptance. Shifting beliefs, contradictory beliefs, borderline belief, alief – these may all have their place to play in explaining human behavior. But when it comes to representational conflict, we need recourse to the notion of acceptance.

To conclude, I will consider some of the implications that recognizing acceptance has for epistemology.

### **Knowledge Without Belief**

Lehrer once presented us with the case of the gypsy lawyer. The gypsy lawyer believed that his client was innocent on the basis of a Tarot card reading. But he then noticed a knock-down argument showing the evidence established his client's innocence. His belief was still (irrationally) based on the cards – defeat the evidence and the lawyer would still believe in his defendant's innocence. We were supposed to intuit that the lawyer knew that his client was innocent. Lehrer concluded that one could know on the basis of the evidence even if one didn't believe on that basis.<sup>73</sup>

Green allows us to draw a more startling conclusion.<sup>74</sup> It seems clear that Green knows that the defendant is innocent. The defendant *is* innocent, the evidence adequately supports this conclusion, Green recognizes the force of the evidence, and he accepts that the defendant is innocent because of the evidence. His willful decision to accept the defendant's innocence is an admirable display of intellectual virtue. And yet, due to irrationality that he is aware of but can't help, Green does not believe that the defendant is innocent. Here we have, I submit, a clear case of knowledge without belief.

---

<sup>73</sup> For an interesting recent discussion of the argument, see (Kvanvig 2003).

<sup>74</sup> One which Lehrer has drawn in some recent, though less widely discussed, work.

Belief is one route to knowledge. I propose that acceptance is another.<sup>75</sup> Whether knowledge requires acceptance is an issue I don't have much space to discuss here, but let me register some doubts. Much of what we believe about our perceptual environment never rises to consciousness in the form of an explicit judgment. But such ordinary beliefs about our environment are paradigm cases of knowledge.<sup>76</sup>

### **Voluntarism and the Ethics of Belief**

Plausibly speaking of belief, Aristotle wrote in the *De Anima* that:

[I]t is plain that having thoughts is not the same as supposing; for the former is up to us, whenever we wish ...; but opining is not up to us – for it is necessary either to hold falsely or to hold truly. (427b16–21)

This is the plausible genesis of a long philosophical tradition of regarding belief as something that is not up to us – something over which we lack direct control. We find the view in Hume and it is, of course, pervasive in modern epistemology. Indeed, the thesis that we lack direct voluntary control over our beliefs is typically taken as a truism in current debates.<sup>77</sup>

---

<sup>75</sup> Of course, one must accept on the basis of evidence in order to know – accepting for moral or prudential reasons won't work.

<sup>76</sup> Lehrer, of course, would disagree. For an illuminating dispute between Lehrer and Dretske on just this topic, see (Dretske 1991) and (Lehrer 1991).

<sup>77</sup> This fact reveals that contemporary epistemologists really have had belief, and not acceptance, in mind.



Of course, there is another distinguished philosophical tradition that regards belief as partly an act of will – we can will to believe. We find this view in the ancient Stoics, in Augustine, James, and (perhaps most famously) in Descartes. The view was typically that belief meant assenting to a proposition; and assent is an act of will. Carl Ginet has recently defended the view that we can, at least sometimes, decide to believe (2001).

I think it is plain that we cannot decide to believe. And if we read the traditional arguments for doxastic voluntarism, they all seem to involve crude mistakes.<sup>78</sup> But charity demands that we treat these thinkers carefully. It is possible that they are not talking about belief at all, but actually have something like acceptance in mind. In the case of Descartes and Ginet, I think one can argue quite forcefully that they are talking about belief and not acceptance.

This allows us to reconcile the two traditions. We need not view this as a great debate over the thesis of doxastic voluntarism. Instead, we have one camp advocating doxastic involuntarism and the other advocating acceptance voluntarism. These are, after all, consistent views.

If this is right, it is very interesting that a host of important philosophers have regarded acceptance as the central propositional attitude in epistemology. This, I think, might plausibly flow from the fact that acceptance, as something under our voluntary

---

<sup>78</sup> For an interesting discussion of the history of doxastic voluntarism, see (Barnes 2006).

control, might ground duties and obligations in a way that belief can't. Consider the Lockean duty to "believe" only what we have adequate evidence for. In a tradition concerned with the voluntary state of acceptance, the thesis is easily intelligible: we are obligated to accept what our evidence supports (and refrain from accepting what it does not) because this is within our control.

In the tradition concerned with the involuntary state of belief, the thesis becomes problematic. How can we have a duty to believe in accord with the evidence if what we believe is not up to us? Of course, much ink has been spilled in recent years trying to answer this question – perhaps the indirect (and spotty) control we have over belief is good enough to ground duties; perhaps "epistemic oughts" don't imply "cans." Most of these accounts are unsatisfying, to my mind. But they are also unnecessary once we recognize that acceptance might be unproblematically governed by duties – moral duties, even.

What we need is not an ethics of belief, but an ethics of acceptance. Such an ethics becomes especially important, I think, as we learn more and more about the recalcitrant irrationality inherent in human belief. But that's a discussion for another day.

## REFERENCES

- Adler, J. (2002). "Akratic Believing?" *Philosophical Studies*, 110, 1–27.
- Audi, R. (1994). "Dispositional Beliefs and Dispositions to Believe." *Nous*, 28(4), 419–434.
- Audi, Robert. (2001). "Doxastic Voluntarism and the Ethics of Belief." In *Knowledge, Truth, and Duty*, ed. Matthias Steup, 93–111. Oxford University Press.
- Barnes, J. (2006). "Belief Is Up to Us." *Proceedings of the Aristotelian Society*, 106(2), 187–204.
- Bennett, Jonathan. (1990). "Why Is Belief Involuntary?" *Analysis*, 50(2), 87–107.
- Blanton, H. & Jaccard, J. (2008). "Unconscious Racism: A Concept in Pursuit of a Measure." *Annu. Rev. Sociol.*, 34, 277–97.
- BonJour, L. (1978). "Can Empirical Knowledge Have a Foundation?" *American Philosophical Quarterly*, 15(1), 1–13.
- Chisholm, Roderick. 1961. "What Is It to Act Upon a Proposition?." *Analysis*, 22(1), 1–6.
- Cohen, L. J. (1989). "Belief and Acceptance." *Mind*, 98(391), 367–389.
- Cummins, Robert. (1996). *Representations, Targets, and Attitudes*. Cambridge, MA: MIT.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.
- DeSousa, R. (1971). "How to Give a Piece of your Mind; or, The Logic of Belief and Assent." *Review of Metaphysics* 25: 52–79.
- . (1988). *Explaining Behavior*. Cambridge, MA: MIT.
- Dretske, F. (1991). "Two Conceptions of Knowledge: Rational vs. Reflective Belief." *Grazer Philosophische Studien* 40: 15–30.
- Elga, A. (2005). "On Overrating Oneself... and Knowing It." *Philosophical Studies*, 123, 115–124.
- Evans, I. (Forthcoming). "The Problem of the Basing Relation." *Synthese*.
- Evans, I. (Unpublished a). "The Dispositionalist Gambit."

- Evans, I. (Unpublished b). "Acceptance, Alief, and In-Between Belief."
- Evans, I. & Smith, N. (2012). *Knowledge*. Malden, MA: Polity Press.
- Feldman, Richard. (2008). "Modest Deontology in Epistemology." *Synthese*, 161, 339–355.
- Foley, R. (1986). "Is it Possible to Have Contradictory Beliefs?" *Midwest Studies in Philosophy X*, 327–355.
- Fodor, Jerry A. (1985). "Fodor's Guide to Mental Representation: the Intelligent Auntie's Vade-Mecum." *Mind*, 94(373), 76–100.
- Frankish, K. (2007). "Deciding to believe again." *Mind*, 116(463), 523–548.
- Gendler, T.S. (2008a). "Alief and Belief," *Journal of Philosophy* 105: 634–663.
- (2008b). "Alief in Action (and Reaction)," *Mind and Language* 23: 552–585.
- Ginet, C. (2001). "Deciding to Believe," in *Knowledge, Truth, and Duties*, ed. by M. Steup.
- Gertler, Brie. (Forthcoming). "Self-Knowledge and the Transparency of Belief," in Anthony Hatzimoysis (ed.), *-Self-Knowledge*. Oxford: Oxford.
- Hume, D. (2007). *A Treatise of Human Nature*, ed. David Fate Norton & Mary J. Norton. New York: Clarendon Press.
- Hunter, D. (2011). "Alienated Belief," *Dialectica*, 65(2), 221–240.
- Hurley, S. (1989). *Natural Reasons*. New York: Oxford University Press.
- Kvanvig, J. (2003). "Justification and Proper Basing," in *The Epistemology of Keith Lehrer*, ed. Erik Olsson. Dordrecht: Kluwer Academic Publishers.
- Lackey, J. (2007). "Norms of Assertion," *Nous* 41.4: 594–626.
- Lehrer, K. (1991). "Reply to Dretske," *Grazer Philosophische Studien* 40: 31–35.
- . (2000a). *Theory of Knowledge*. Westview Press.
- . (2000b). "Acceptance and Belief Revisited," in *Believing and Accepting*, ed. Pascal Engel. Dordrecht: Kluwer Academic Publishers.

- . (2003). “Coherence, Circularity, and Consistency: Lehrer Replies,” in *The Epistemology of Keith Lehrer*, ed. Erik Olsson. Dordrecht: Kluwer Academic Publishers.
- Levy, N. (2004). “Epistemic Akrasia and the Subsumption of Evidence: A Reconsideration,” *Croatian Journal of Philosophy*, 4, 149–56.
- Lycan, W.G. (2008). “Phenomenal Intentionalities,” *American Philosophical Quarterly* 45: 233–52.
- Mele, A. (1987). *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. Oxford: Oxford University Press.
- Owens, D. (2002). “Epistemic Akrasia.” *The Monist*, 85(3), 381–97.
- Petit, P. & Smith, M. (1996). “Freedom in Belief and Desire,” *Journal of Philosophy*, 93, 429–77.
- Ribeiro, B. (2011). “Epistemic Akrasia.” *International Journal for the Study of Skepticism*, 1, 18–25.
- Rowbottom, D.P. (2007). “In-Between Believing’ and Degrees of Belief,” *Teorema* 26: 131–137.
- Ryle, G. (2009). *The Concept of Mind*. New York: Routledge.
- Schwitzgebel, Eric. (2002). “A Phenomenal, Dispositional Account of Belief.” *Nous*, 36(2): 249–275.
- Setiya, K. (2008). “Believing at Will.” *Midwest Studies in Philosophy*, 32, 36–52.
- Shoemaker, S. (1996). *The First-Person Perspective and Other Essays*. New York: Cambridge University Press.
- Silins, N. (2012). “Judgment as a Guide to Belief.” *Consciousness and Introspection*. Stoljan, Daniel and Smithies, Declan (eds.), Oxford University Press.
- Sommers, Fred. (2009). “Dissonant beliefs,” *Analysis* 69: 267–274.
- Steup, Matthias. (2008). “Doxastic Freedom.” *Synthese*, 161, 375–392.
- Titelbaum, M. (Unpublished). “In Defense of Right Reason.”

- van Fraassen, C. (1984). "Belief and the Will." *The Journal of Philosophy*, 81(5), 235–256.
- Weatherson, Brian. (2008). "Deontology and Descartes'S Demon." *Journal of Philosophy*, 105(9), 540–569.
- Williamson, Timothy. (2000). *Knowledge and Its Limits*. New York: Oxford University Press.
- Wilson TD, Lindsey S, Schooler TY. (2000). "A model of dual attitudes." *Psychol. Rev.* 107: 101–26.
- Winters, Barbara. (1979). "Believing at Will." *The Journal of Philosophy*, 76(5), 243–256.
- Zimmerman, A. (2007). "The Nature of Belief," *Journal of Consciousness Studies* 14.1: 61–82.