

The Sources of Epistemic Normativity

By

Clinton Castro

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

(Philosophy)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 4/24/2018

The dissertation is approved by the following members of the Final Oral Committee

Michael Titelbaum, Associate Professor, Philosophy

Sarah Paul, Associate Professor, Philosophy

Alan Sidelle, Professor, Philosophy

Jesse Steinberg, Faculty Associate and Assistant to the Chair, Philosophy

Alan Rubel, Associate Professor, The Information School

*For Molly and Ray.*

# Acknowledgements

I owe a great number of people a great deal of gratitude for this dissertation. I don't know how to properly thank them all. Not knowing how else organize them, they're thanked (roughly) in order of appearance in my life.

First and foremost, thanks to my mother for her love and blind faith in my abilities. Without her support I would not have completed this project. Thanks to my brother for his contrarianism and pessimism. Without him, this project would not be as good as it is. Thanks to my first grade teacher, Debbie Graham, for teaching me to be intellectually curious. Thanks to my 6-8<sup>th</sup> math grade teacher, Nancy Berkeland, for teaching me the importance of rigor. Thanks to my 12<sup>th</sup> grade English teacher, Mary Beth Carosello, for introducing me to philosophy.

I am indebted to a group of friends from high school with whom I took most of my classes on account of being in advanced math at a very small high school. Many thanks to Christine and Joyce Ogasawara, Brian and Chistina Yamada, Arun Menon, Cassidy Morris, Matthew Honea, and Kenny White. (But especially Arun.) If it wasn't for your positive influence, I likely wouldn't be an academic.

I owe tremendous thanks to the Cal State University, Long Beach philosophy faculty. Special thanks are owed to Wayne Wright, Nellie Wieland, Cory Wright, Marcy Lascano, Alex Klein, Larry Nolan, and Jason Raibley. Thanks to Trevor Elder for keeping philosophy fun. Thanks to the CSULB McNair scholars Program—and

Dr. Howard Wray, especially—for making graduate school a serious possibility for me.

Thanks to my dogs, Henry and Olive, for weaseling their way into my life when I was in no position to be adopting animals. Their companionship is invaluable. I probably wouldn't have made it through graduate school without their love and support.

To the entire UW-Madison faculty and staff, thank you for being such a wonderful and welcoming community. Many special thanks are owed here. To Patty Winspur, Lori Grant, Christy Horstmeyer, Nina Akli, Cheryl Schutte, and Miriam Fagan, thank you so much for your support, patience, and kindness. To Alan Sidelle, Sarah Paul, and Michael Titelbaum, thank you for feedback on this project and for your mentorship. To Alan Rubel and Jesse Steinberg, thank you for serving on my committee. To Hayley Clatterbuck, Hadley Cooney, Casey Hart, Adam Pham, Brian McLoone, Tim Aylsworth, Olav Vassend, David O'Brien, Zi Lin, Josh Mund, Emi Okayasu, Ben Schwan, Shanna Slank, and Reuben Stern: I'm so appreciative of the philosophical conversations we've had.

I owe more thanks than I can give to my wife, Molly. I'll never understand how I got so lucky.

Last but not least, thanks to my son, Ray. You mostly made finishing this dissertation more difficult, but you also made finishing it worthwhile.

# Abstract

Belief is beholden to a norm:

(T) A belief that  $p$  is correct only if it's the case that  $p$ .

Not only are beliefs judged against this norm; beliefs are also regulated by it. You can't, try as you might, intentionally form beliefs contra (T), even when you seem to have excellent reason to. You can't, for example, bring yourself to believe that  $2 + 2 = 5$  for any amount of money. This is puzzling—why is it that in the case of belief, what is correct is so closely tied to what we can do?

In this dissertation I defend an intuitive account of belief that provides an easy answer to this question:

**Teleology:** part of what it is for something to be a belief is for it to be guided by a truth-oriented goal, the satisfaction of which would involve satisfying (T).

Simply put, Teleology understands the activity of forming a belief as the activity of trying to form a true belief. In understanding this orientation to the truth as part of the nature of belief, Teleology tidily explains why it's incoherent to intentionally form a false belief: to intend to believe falsely is to intend to believe both truly and falsely. In providing this answer to our question, Teleology also explains (T) itself: since beliefs just are attempts to believe truly, a belief is defective if it isn't true.

Recently, there has been an uptick in interest in the nature and source of epistemic norms. Teleology offers a promising strategy for approaching these issues. It locates the source of a fundamental norm of belief in a familiar fact about belief. Despite its initial appeal, however, the teleological approach has been underappreciated. This is because Teleology is thought to face three serious challenges. I organize my dissertation around my responses to these challenges.

The first challenge is one that any goal-oriented account of epistemic normativity has to face. It's a truism that if you have evidence that  $p$ , you thereby have a reason to believe  $p$ . But if epistemic norms apply to us in virtue of our engaging in some goal-oriented practice, it seems that there will be cases where we have evidence that  $p$  but no goal that would be served by believing in accord with our evidence. I address this challenge by first showing that there are two ways of understanding the force of epistemic norms: either they are genuinely normative, as moral norms are often thought to be, or, like the norms of etiquette or games, they lack such normative force. I argue against the claim that epistemic norms are genuinely normative, and show that teleologists can account for the truism once we accept that evidential reasons lack genuine normative force.

The second challenge is that of explaining the significance of epistemic justification. Beliefs can be defective, even when they are true. More specifically,

(J) A belief that  $p$  is correct only if it's justified.

How can the teleologist account for this? After arguing against attempts to derive (J) from (T), I argue that beliefs aim at the truth but not only the truth. On the account I argue for, beliefs aim to be non-accidentally true. To be *non-accidentally* true, a belief that  $p$  must not only itself be true; it must stand in a stronger relationship to  $p$ . This is how we get (J).

The third challenge is that of explaining the exclusive role that evidence plays in first-person deliberation about what to believe. When we deliberate about what to believe, we only consider what we take to be the evidence. Call this the phenomenon of *transparency*. Teleologists, it seems, could account for transparency by saying that it's part of the nature of belief to be solely sensitive to truth-related considerations. But if this is the case, then certain mental states that seem to be beliefs will not—per Teleology—count as beliefs. The products of wishful thinking, for example, seem to be beliefs, but wishful thinking is not solely sensitive to truth-related considerations. I respond by arguing that the phenomenon of transparency is not what it has been taken to be. In deliberation we don't always consider the evidence. Rather, we consider what we *take* the evidence to be. Once this is understood, teleologists can advocate for an account where it's part of the nature of belief to be only somewhat sensitive to truth-related considerations. Thus, the teleologist can explain transparency and yet consider wishful thoughts beliefs.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The Teleological Conception of Belief: A Preliminary Case . . . . .	1
1.2	So, Why Aren't We All Teleologists? . . . . .	3
1.3	In defense of Teleology . . . . .	5
<b>2</b>	<b>Evidence for Instrumentalism</b>	<b>7</b>
2.1	Introduction . . . . .	7
2.2	Enough Reasons? . . . . .	12
2.2.1	Special Interest Accounts . . . . .	12
2.2.1.1	The Aim of $\phi$ -ing for Reasons . . . . .	13
2.2.1.2	Truth as the Aim of Belief Forming Processes . . . . .	16
2.2.1.3	General Remarks . . . . .	17
2.2.2	Modal Accounts . . . . .	17
2.2.2.1	Hypotheticalism . . . . .	18
2.2.2.2	Rule Instrumentalism, Modalized . . . . .	20
2.2.2.3	General Remarks . . . . .	21
2.3	The Compelling Argument, Revisited . . . . .	23
2.3.1	Evidence Always Provides a Genuine Reason? . . . . .	26
2.3.2	Too Few Normative Reasons? . . . . .	27
2.3.2.1	The Teleological Solution . . . . .	29
2.3.2.2	The Normativist Solution . . . . .	30
2.3.3	Genuinely Normative (Non)evidential Reasons? . . . . .	32
2.4	Conclusion . . . . .	34
<b>3</b>	<b>Why Justify Belief?</b>	<b>36</b>
3.1	Introduction . . . . .	36
3.2	Instrumentalism . . . . .	39
3.3	Epistemic Consequentialism . . . . .	41
3.3.1	Direct Epistemic Consequentialism . . . . .	41
3.3.2	Indirect Epistemic Consequentialism . . . . .	44
3.4	Epistemic Teleology . . . . .	49



3.4.1	Completing Step 2 . . . . .	51
3.4.1.1	Requirement N? . . . . .	51
3.4.1.2	Requirement B? . . . . .	52
3.4.1.3	Requirement R? . . . . .	52
3.4.2	The Aim of Guessing . . . . .	53
3.5	What Went Wrong? . . . . .	56
3.6	Safely Pursuing the Truth . . . . .	57
3.7	Applications . . . . .	60
3.7.1	Evidentialism . . . . .	60
3.7.2	Process Reliabilism . . . . .	61
3.7.3	Evidentialism and Reliabilism . . . . .	63
3.8	Conclusion . . . . .	64
<b>4</b>	<b>Transparent Teleology</b> . . . . .	<b>65</b>
4.1	Introduction . . . . .	65
4.2	Belief's Normativity via the Aim of Belief . . . . .	68
4.2.1	Teleology and the Aim of Belief . . . . .	68
4.2.2	Normativism and the Aim of Belief . . . . .	70
4.3	The Teleologist's Dilemma . . . . .	71
4.4	Defusing the Dilemma . . . . .	72
4.4.1	Against (10) . . . . .	73
4.4.2	Against (5) . . . . .	76
4.4.2.1	Varieties of Strong Truth-Regulation . . . . .	76
4.4.2.2	Strong Truth-Regulation, Cognitive Bias, and the Shotgun Approach . . . . .	79
4.4.2.2.1	<i>U</i> & Biased Thinking . . . . .	80
4.4.2.2.2	Formally Rational Standards & Wishful Think- ing . . . . .	82
4.4.2.3	Earnest Misapplication of Correct Standards . . . . .	84
4.4.2.4	Earnest Misapplication of Incorrect Standards . . . . .	85
4.4.2.5	General Remarks . . . . .	86
4.4.3	Against (7) . . . . .	87
4.4.3.1	Weak Truth Regulation and Transparency . . . . .	87
4.5	Conclusion . . . . .	88
	<b>Bibliography</b> . . . . .	<b>90</b>

*You must learn to live with your own conscience  
Your own morality  
Your own decision  
Your own self!*

*You alone can do it: There is no authority but yourself.*

*—Crass, Yes sir, I will*

*Believe Truth! Shun Error!*

*—William James, The Will to Believe*

# Chapter 1

## Introduction

### 1.1 The Teleological Conception of Belief:

#### A Preliminary Case

Beliefs are subject to certain normative standards. Here are two statements those standards imply:<sup>1</sup>

(T) There's something wrong with false beliefs.

(J) There's something wrong with unjustified beliefs.

Why is it that beliefs are beholden to norms of justification and truth?

Here's a hypothesis:

What distinguishes believing a proposition from imagining or supposing it is [...] the aim of getting the truth value of that particular proposition right, by regarding the proposition as true only if it really is (Velleman, p. 252)

Call this the *teleological conception of belief*. This hypothesis is attractive in part because it can explain (T) and (J). (T) comes out true because if all beliefs are held

---

<sup>1</sup>I identify statements the standards imply as opposed to the standards themselves because there is too much controversy over the proper formulation of the standards themselves.

with the aim that those beliefs are true, then false beliefs are defective by the lights of their own constitutive aim. Further, it's plausible that if you aim to believe truly, then rationality compels you to have justified beliefs. If this is correct, then believing unjustifiably involves the violation of a norm.

So, the teleological conception plausibly explains (T) and (J). But, so does the following *objective hypothesis*: beliefs ought to be true. (T) follows immediately from this hypothesis, and (J) isn't far behind. Suppose I ought to believe truly. Now make the plausible assumption that if I ought to  $\phi$ , then I ought do what I rationally expect amounts to  $\phi$ 'ing. Given the assumption that the teleologist helped herself to in her account—i.e., that if you aim to believe truly, then rationality compels you to have justified beliefs—the objective hypothesis also explains (J).

So, both hypotheses explain (T) and (J). Do we have any reason to prefer one to the other? To answer this question, let's make a few more observations about belief.<sup>2</sup>

**It is difficult or impossible to believe at will.** Let's say that if I  $\phi$  that  $p$  without any regard for  $p$ 's truth, I  $\phi$  *at will*. We seem unable to believe at will. But this is not true of other attitudes; for example, I can imagine that  $p$  regardless of whether I think  $p$  is true.

**It's absurd to have Moorean beliefs.** There is something essentially surd about thinking to oneself, "I believe  $p$ , but not- $p$ ." But again, this is not true for many attitudes: there's nothing wrong thinking "I assume  $p$ , but not- $p$ ."

**Evidential considerations have special force in deliberative contexts.** When I ask myself *whether to believe  $p$* , my question is settled by the answer to the question *whether  $p$  is true*. But, the questions *whether to believe  $p$*  and *whether  $p$  is true* are facially different questions—why should they invariably be settled by the same considerations?

The teleological conception of belief can explain these phenomena. On the hypothesis, we shouldn't be able to believe at will, since part of what it is for an attitude to

---

<sup>2</sup>I owe the items on this list to Fassio (2015).

be a belief is for it to be concerned with the truth; according to teleologists, to form an attitude with complete indifference to the truth is to form a non-belief. Similarly, the absurdity of Moorean beliefs is explained because believing  $p$  involves regarding  $p$  as true because it really is. And lastly, if belief's *raison d'être* is truth, then we should expect *whether I should believe  $p$* , to be settled by the answer to the question *whether  $p$  is true*.

None of these observations are explained by the objective hypothesis. It cannot explain the difficulty of believing at will. After all, we are fully capable of doing what we oughtn't. For example, it's plausible that

**The Golden Rule:** we ought to treat others only as we consent to being treated in the same situation.<sup>3</sup>

But I don't always live up to this ideal, unfortunately. Similarly, that I *ought* to believe the truth doesn't explain the absurdity of Moorean thoughts or the special force of evidential considerations in deliberative contexts.

## 1.2 So, Why Aren't We All Teleologists?

The above constitutes a compelling *prima facie* case for teleology. But not everyone accepts the teleological conception of belief. Why not? Here I will discuss two related reasons.

The first reason is the presence of attractive alternatives. Consider

**The Normativist Conception of Belief:** It's part of the concept 'belief' that <necessarily, for all  $p$ , you ought to believe that  $p$  only if  $p$ ><sup>45</sup>

---

<sup>3</sup>For a defense see Gensler (2013).

<sup>4</sup>This formulation of the truth-norm is owed to Gibbons (2013)

<sup>5</sup>This is not the only formulation of the view. There's quite a lot of debate among normativists over what, exactly, the proper formulation of the norm is.

The normativist might say that because <necessarily, for all  $p$ , you ought to believe that  $p$  only if  $p$ > is part of the concept 'belief', possessing the concept 'belief' involves the acceptance of a prescription, like <to believe that  $p$  only if  $p$ > (viz. Shah (2003)). One argument for normativism is that it can explain all of our observations, and it can do so *better* than the teleological conception can.<sup>6</sup>

The second reason is that the teleological conception of belief has been subject to three influential objections:

**Too Few Reasons.** The teleological conception is often paired with reasons-instrumentalism: the claim our reasons depend on our ends. Further, there are some propositions about which some agents have no interest in believing truly. However, epistemic norms are categorical: if one has evidence that  $p$ , one is thereby in possession of a reason to believe  $p$ . The teleologist, it seems, cannot explain this if she is committed to reasons-instrumentalism. (cf. Kelly (2003))

**The Problem of Guessing.** Belief is not the only attitude characterized by the aim of getting the truth value of the particular proposition it is directed towards right; guessing can also be characterized this way. But ( $J_g$ ), the claim that there's something wrong with unjustified guesses, is not true. So, we should be skeptical that the teleological conception of belief explains ( $J$ ). (cf. Owens (2003))

**The Teleologist's Dilemma.** The teleologist can explain the special force of evidential considerations in deliberative contexts only if she incurs a high cost. She can explain it only if the aim of belief is characterized by a high degree of sensitivity to truth-related considerations. But if this is the case, then certain mental states that seem to be beliefs will not—per Teleology—count as beliefs. The products of wishful thinking, for example, seem to be beliefs, but wishful thinking is not highly sensitive to truth-related considerations. (cf. Shah (2003))

This dissertation is a general defense of the teleological conception of belief, and it is organized around my responses to these three challenges.

---

<sup>6</sup>Of course, I do not think that this claim is true. My arguments are presented in Chapter 4.

### 1.3 In defense of Teleology

In Chapter 2, “Evidence for Instrumentalism”, I respond to the Too Few Reasons problem. I address this challenge by first showing that there are two ways of understanding the force of epistemic norms: either they are genuinely normative, as moral norms are often thought to be, or, like the norms of etiquette or games, they lack such normative force. I argue against the claim that epistemic norms are genuinely normative, and show that teleologists can account for the truism once we accept that evidential reasons lack genuine normative force.

In chapter 3, “Why Justify Belief?”, I respond to The Problem of Guessing. I argue that beliefs aim at the truth but not only the truth. On the account I argue for, beliefs aim to be non-accidentally true. To be *non-accidentally* true, a belief that  $p$  must not only itself be true; it must stand in a stronger relationship to  $p$ . Endeavoring to have one’s belief be non-accidentally true, I argue, is what grounds (J) and distinguishes the activity of forming a belief from the activity of guessing.

In chapter 4, “Transparent Teleology”, I respond to The Teleologist’s Dilemma. I respond by arguing that the special force of evidential considerations in deliberative contexts has been misunderstood. In deliberation we don’t always consider the evidence. Rather, we consider what we *take* the evidence to be. Once this is understood, teleologists can advocate for an account where it’s part of the nature of belief to be only somewhat sensitive to truth-related considerations. Thus, the teleologist can explain transparency and yet consider wishful thoughts beliefs.

In responding to these challenges, I not only defend the teleological conception of belief from these influential objections. I also show that it’s not the case that normativism explain all of our observations better than the teleological conception can.

In so doing, I demonstrate that teleology is worthy of more serious consideration than it has received.



## Chapter 2

# Evidence for Instrumentalism

**Abstract.** It seems that if you have evidence that  $p$ , you have a genuine reason to believe  $p$ . Why? According to normative instrumentalism, it's because believing the truth is useful, relative to your ends. But can this view really work? After all, it's commonly thought that evidence that  $p$  is *always* a genuine reason to believe  $p$ . And, believing the truth is sometimes entirely useless. Some instrumentalists respond to this challenge by arguing that it is, in fact, always useful to believe the truth. This, I argue, is a mistake. Instead, instrumentalists must respond to the challenge by denying the commonplace that evidence that  $p$  always provides a genuine reason to believe  $p$ . I argue that this commonplace is false and explore some of the implications of denying it.

### 2.1 Introduction

According to a family of attractive metanormative theories, your genuine reasons depend on your ends. I say “genuine” because metanormativity theorists are interested in a particular sense of “reason,” the sense which calls to mind the idea of a standard we are robustly criticizable for not meeting.<sup>1</sup> I mean to be inclusive about the meaning of “ends”—the term is meant to include interests, desires, commitments, convictions, and so on—so long as those ends are in some way derived psychologically. Which ends are normatively relevant varies from

---

<sup>1</sup>See Maguire and Woods (*ms.*) for this characterization of genuine normativity.

person to person.

There is, of course, controversy over what characteristics genuine reasons have, but some general remarks can be made. One generally accepted characteristic is that genuine reasons motivate us to do what they favor. This is one place where the attractive family of theories, call it *normative instrumentalism*, has an edge: in grounding agents' reasons in their ends, it is in good position to explain this phenomenon.

Another attraction of normative instrumentalism is its promise to naturalize normativity. For wide swaths of philosophers, *naturalism*, which I'll understand as the metaphysical thesis that "all real properties are those that would figure ineliminably in perfected versions of the natural and social sciences" is dogma (Shafer-Landau (2003)). Normative instrumentalists understand genuine reasons as constituted by perfectly naturalistically respectable entities like belief-desire pairs, and so normative instrumentalists are in a good position to give an account of normativity that comports with a naturalistic world-view.

Normative instrumentalism is not without its drawbacks, of course. It is not difficult to think of cases where, intuitively, an individual has a genuine reason to perform some action,  $\phi$ , and yet has no end that speaks in favor of  $\phi$ -ing. Consider, for example, an anorexic who "accepts norms that prescribe death by starvation, if the alternative is a figure plump enough to sustain life"(Gibbard (1990) p. 171). Or, perhaps, an abusive husband who does not care about his partner's well-being (Williams (1979)). In these cases we feel that the anorexic has a genuine reason to consume enough calories to stay alive, and that the husband has a genuine reason to stop the abuse. Normative instrumentalism, however, is in a poor position to vindicate these intuitions.

I don't know what normative instrumentalists should say about cases like the

above, where agents seem to have genuine moral or prudential reasons and yet have no end-given reasons to underwrite them. Normative instrumentalists might be able to demonstrate that the characters in stories like these actually do have ends that speak in favor of eating and kindness, that—for example—given a sufficiently weak understanding of what it takes to promote a desire, the husband has at least *a* reason to choose kindness (Schroeder, 2007). Or, they may be able to talk us out of thinking that these characters have the reasons that we want to attribute to them. Once we *carefully* imagine the case of the anorexic, the normative instrumentalist might say, we'll see that she has no reason—prudential or otherwise—to eat enough to sustain her life (Street, 2009).

In any event, there is a domain of reasons that seems particularly resistant to either of these strategies. In the “theoretical” or “epistemic” domain, we speak of evidence as our reason for having a particular doxastic attitude towards a certain proposition. Clearly, our evidence can provide reasons for adopting attitudes that have nothing to do with our interests. Nevertheless, it seems that evidence that *p* always provides a genuine reason for believing *p*. This spells a special kind of trouble for the normative instrumentalist.

Consider Tom Kelly's (2003) remarks on *epistemic instrumentalism*, the view that our epistemic reasons depend on our ends:

Whether Bertrand Russell was right- or left-handed, whether Hubert Humphrey was an only child—these are matters of complete indifference to me. That is, I have no preference for having true beliefs to having no beliefs about these subjects; nor, for that matter, do I have any preference for having true beliefs to false beliefs. There is simply no goal—cognitive or otherwise—which I actually have, which would be better achieved in virtue of my believing true propositions about such subjects, or which would be worse achieved in virtue of my believing false propositions about them. However, from the fact that some subjects are matters of complete indifference to me, it does not follow that I will inevitably lack epistemic reasons for holding beliefs about those

subjects (Kelly (2003), p. 624).

In this passage, Kelly lays the foundation for the following argument:

### **The Compelling Argument**

*Instrumentalists  
Have Too Few  
Reasons*      If epistemic instrumentalism is true, then it's not the case that if an agent has evidence that Bertrand Russell is left-handed, she thereby has a genuine reason to believe that he was left-handed.

*Evidence Always  
Provides a  
Reason*      If an agent has evidence that Bertrand Russell is left-handed, she thereby has a genuine reason to believe that he was left-handed.

So, epistemic instrumentalism is false.

This argument poses a serious challenge for epistemic instrumentalists. Since normative instrumentalism entails epistemic instrumentalism, it also poses a serious challenge for normative instrumentalists.

The first premise of The Compelling Argument only relies on the plausible conjecture that there are some *exception cases*, cases where an agent's evidence favors believing  $p$ , but her ends do not provide a reason to believe in accordance with the evidence with respect to  $p$ . If you have trouble imagining how facts about Bertrand Russell's handedness could *possibly* be irrelevant to anyone's ends, you only have to think of something that seems more irrelevant, maybe whether *I* am left-handed, or whether my father was. Whether the conjecture that instrumentalists have too few reasons pans out will—of course—turn out to be a matter of debate, as several epistemic instrumentalists argue that each agent has *some* end that is promoted by believing in accord with her evidence. I will offer reasons to think that this strategy for responding to The Compelling Argument cannot work.

The second premise, "Evidence Always Provides a Reason," relies on the plausible thought that evidence that  $p$  always provides a genuine reason to believe  $p$ .

This, like the first premise, seems on its face undeniable. As a small bit of evidence for this claim, consider the awkwardness of claims like “I have evidence that it’s raining, but no good reason to believe it.”<sup>2</sup> Such statements sound confused. This, I take it, is evidence that evidence that *p* just is a genuine reason to believe *p*.

Appearances to the contrary, I think normative and epistemic instrumentalists can give a satisfactory response to The Compelling Argument. But to do this they must take a controversial approach, one that denies “Evidence Always Provides a Reason.” This is because there is no plausible way to deny “Instrumentalists Have too Few Reasons.” A large portion of this chapter is dedicated establishing this. The key to rejecting “Evidence Always Provides a Reason,” I will argue, is seeing that evidence that *p* does always provide a reason to believe *p*, it just doesn’t always provide a *genuine* reason to believe *p*.<sup>3</sup> If we accept this, the instrumentalist is still left with the challenge of explaining how her position isn’t undermined by an updated version of Kelly’s argument, one that argues from the claims that instrumentalists have too few (non-genuine) reasons and evidence always provides a (non-genuine) reason, to the conclusion that epistemic instrumentalism is false. But this—as I will demonstrate—is a challenge to which instrumentalists can rise.

---

<sup>2</sup>I use “good” instead of “genuine” because “genuine” sounds a bit artificial. It’s hard to imagine an ordinary case where “genuine”(as opposed to “good”) is the natural thing to say. I think the substitution is innocuous. I take it to be very plausible that in this context “good” and “genuine” have the same meaning.

<sup>3</sup>While “Evidence Always Provides a Reason” is the standard view, I am not alone in rejecting it. Hazlett (2013), Papineau (2013), and Maguire and Woods (*manuscript*) also hold that epistemic reasons are not genuinely normative.

## 2.2 Enough Reasons?

It is helpful to sort instrumentalist rejections of “Instrumentalists Have Too Few Reasons” into two categories.<sup>4</sup> One strategy, the *special interest approach* argues that there is some end that agents have in common that is promoted by believing in accord with evidence. The other strategy, the *modal approach*, argues that apportioning one’s beliefs to the evidence is uniquely and universally useful—that no matter what your interests are, believing on your evidence will promote those interests. I’ll discuss these strategies in turn.

### 2.2.1 Special Interest Accounts

One might try to defend epistemic instrumentalism with the conjecture that there is some cognitive goal common to all agents that is promoted by believing the truth or believing in accord with the evidence. The interest could be *now* believing the truth (Foley, 1987), correctly accurately representing the world (BonJour, 1985), or maximizing truth and minimizing falsity in a large body of beliefs (Alston, 2005). The special interest approach is often quickly dismissed on the grounds of The Compelling Argument: it just isn’t plausible that all agents actually have any of these cognitive goals. Examples like Kelly’s show that there are some propositions that some agents have *no* interest in. This thought often leads to a dismissal of any account that proceeds by way of the special interest strategy. I ultimately agree that special interest accounts are flawed for the reason Kelly cites. But I think that a rejection of this strategy from reflection on this naïve version of the view is too quick, as there are aims that might be constitutive of agency or belief that saddle us with a concern for the truth that is not immediately apparent.

---

<sup>4</sup>I owe this terminology to Sharadin (*forthcoming*).

### 2.2.1.1 The Aim of $\phi$ -ing for Reasons

Charles Côté-Bouchard (2015) responds to “Instrumentalists Have too Few Reasons” by noting that agents essentially aim to  $\phi$  for reasons, where  $\phi$  is anything we do or attitude we have. The idea is that for any arbitrary instance where you  $\phi$ , you  $\phi$  because you believe that there is a reason,  $p$ , that speaks in favor of  $\phi$ 'ing. When you  $\phi$  and take  $p$  to be your reason for  $\phi$ 'ing, you aim for your belief to be true: you want it to be the case that you did not  $\phi$  in error.

Let's grant this. It does not yet follow that we have a reason to believe  $p$  if our evidence supports  $p$ . This is because there are cases where believing in accord with my evidence won't promote the end of  $\phi$ 'ing for reasons. The most obvious case is one like Kelly's: if I'll *never* act on information relating to Russell's handedness and don't care about the issue for its own sake, I don't—on this account—have a reason to take an attitude towards the proposition that Russell was left-handed, let alone the attitude my evidence supports.

Recognizing this, Côté-Bouchard endorses what I'll call *rule instrumentalism*. Rather than say an agent has a genuine reason to  $\phi$  if and only if so doing would promote her ends, the rule instrumentalist says one has a reason to  $\phi$  if  $\phi$ -ing is favored by some general rule the agent should adopt, given her ends. Côté-Bouchard is not alone in turning to rule instrumentalism to address this kind of difficulty. Both Kornblith (1993) and Leite (2007), for example, make this move in defense of their brand of epistemic instrumentalism (which will receive its own treatment in the next section.) In what follows, I hope to make an argument against any view that appeals to rule instrumentalism in response to the too few reasons challenge.

Rule instrumentalists are motivated to accept their view because the act-based

version of their view admits of exception cases.<sup>5</sup> Herein lies the difficulty for any rule instrumentalist view. For any rule instrumentalist view, we can imagine an agent who is motivated to follow the act-based version of that view, is capable of adhering to that view, and knows she is in an exception case. Such an agent will pose a threat to the rule instrumentalist: she will lack any end-given motivation to believe in accord with the evidence in the exception case we imagine her to be in. This shows that the rule instrumentalist response to “Instrumentalists Have Too Few Reasons” comes at a high cost: it gives up on the intuitive thought that our genuine reasons should speak to the ends that motivate us.

The rule instrumentalist may reply by claiming that the above argument begs the question: it assumes an account of what is involved in promoting an end, and argues against the rule instrumentalist account of end-promotion from that very assumption. I don't think this is quite right, however. The instrumentalist position is attractive because *on an intuitive picture of what is involved in promoting an end*, a claim like, “we have a reason to  $\phi$  when  $\phi$ -ing promotes our ends” sounds very plausible. The root of the problem that I am identifying is that the rule instrumentalist picture of what it takes to promote an end is far from intuitive.

At this point, it is worth discussing the extent to which my criticism of rule instrumentalism resembles the “rule worship” objection to rule utilitarianism, and why it is in fact different from this objection in an important way. Rule utilitarians opt for a theory that evaluates actions indirectly, and this is often for reasons that mirror the rule instrumentalist's motivations for taking a rule-based approach. Taking a rule-based approach, of course, leads rule utilitarians to prohibit some ac-

---

<sup>5</sup>The act-based version of a rule instrumentalist view, we'll say, is a view that has the same general features as its rule instrumentalist counterpart—it agrees which ends are normatively relevant, for example—but judges  $\phi$ -ing's directly (as opposed to judging them indirectly via the deployment of general rules).



tions that are optimific and to require some actions that are not. In the extreme case, rule utilitarians require agents to act in ways that will lead to disastrous outcomes, even in the presence of alternatives that will lead to non-disastrous outcomes. The rule worship objection, then, states that there is some kind incoherence in this: utilitarians oughtn't recommend acting in ways that will have disastrous outcomes when disaster can be avoided, as this is incompatible with a commitment to the promotion of overall utility.

The rule-worship objection has been responded to in a number of ways. Most relevant for our purposes is an approach that holds fast to its rules in the face of the charge of rule worship, saying that there is no incoherence involved in requiring suboptimal actions.<sup>6</sup> The response to the charge is that the objector forgets what project the rule utilitarian is engaged in. The rule utilitarian is in the business of choosing optimific *rules* and judging actions against those rules. When the rules require actions that are not optimific she recommends them nonetheless, and there is no incoherence in this.

The rule instrumentalist cannot co-opt this solution. The reason is that the rule instrumentalist is attempting to give a theory of reasons, one that is made plausible—in part—by its ability to explain the motivational force of genuine reasons. An exceptionless form of rule instrumentalism cannot maintain an attractive story of motivation. To drive this point home, consider again the instrumentalist who says that we can capture “Evidence Always Provides a Reason” by ascribing to all agents the goal of now believing  $p$  if and only if  $p$ . This view is unavailable to the instrumentalist because it's implausible that all agents fetishize the truth in

---

<sup>6</sup>There is a more popular response. But it is less relevant to our purposes, as it is not one that will appeal to rule instrumentalists. The response I have in mind is one that makes exceptions to rules in cases where the rules recommend a truly disastrous course of action (Brandt, 1992). This solution, however, is useless to rule instrumentalists in this context because it would amount to a denial of “Evidence Always Provides a Reason.”

this way, and thus any reasons such a view ascribes to them will be alien to them. My contention is that in exception cases, reasons that flow from rules are similarly alien. Rule utilitarianism is not saddled with this constraint. After all, according to rule utilitarianism agents needn't be motivated by the principle of utility, and—for that matter—it might be preferable for them to believe it to be false (Sidgwick, 1901).

### 2.2.1.2 Truth as the Aim of Belief Forming Processes

With the naïve special interests approaches and the  $\phi$ -ing for reasons approach off the table, where else might the special interest theorist turn? The last special interest theory that I will discuss locates epistemic normativity in the nature of belief.

David Velleman (2000) understands the aim of accepting  $p$  if and only if  $p$  as constitutive of the nature of belief and the heart of epistemic normativity. This may sound like a presentation of the naïve thought that agents have the personal goal of, say, believing  $p$  if and only if  $p$ . But, importantly, on Velleman's account, having a personally held goal that  $x$  be the case is only *one* way to have that goal. Another way to have a goal is to have a subpersonal process whose constitutive function or aim is to see to it that  $x$  obtains. Just like it may be said that the constitutive aim of the heart is to pump blood, the constitutive aim of our cognitive faculties is to accept  $p$  if and only if  $p$ .

At the beginning of this essay, I noted that one of instrumentalism's selling points is its promise to make sense of the fact that our genuine reasons are capable of motivating us. One intuitive thought that underwrites instrumentalism is the simple idea that if the reasons an agent has are rooted in her ends, then we can explain how that agent's reasons motivate her. If we now learn that in addition

to having an end in the normal way—whatever that turns out to be—we can also have a desire in virtue of a subpersonal aim or subpersonal processes, there is tension with this fundamental idea. The problem here is that the ends that I have in the normal way are connected to what *I* can be motivated to do in a way that my subpersonal processes' ends are not, unless I take their ends as my own or see their proper functioning as valuable. But if this is the case, the too few reasons problem reemerges. So long as there are agents who are alienated from the aims of their subpersonal processes, we'll have agents who have too few epistemic reasons.

### **2.2.1.3 General Remarks**

Let's briefly take stock. We have considered three attempts to make special interest responses to "Instrumentalists Have too Few Reasons." The naïve version of the approach fails because it's just implausible that agents have the goal it ascribes to them. The central problem with sophisticated versions of this view is the same as the naïve version: agents don't have in common any one particular end that could be used to give instrumentalist grounding for "Evidence Always Provides a Reason." All of this, perhaps, is an artifact of trying to trace evidentialism back to *one* goal. With these thoughts in mind, let's turn to the modal approach.

## **2.2.2 Modal Accounts**

The modal strategy, unlike the special interest strategy, holds that evidence provides a reason, no matter what your interests are. The intuitive thought behind the view is that believing the truth is very helpful in terms of getting what you want. This, I think, is undeniable. The question for the proponent of the modal strategy is, of course, whether this idea can be expanded upon and used as grounds for a

categorical imperative to respect the evidence.

### 2.2.2.1 Hypotheticalism

In *Slaves of the Passions*, Mark Schroeder provides an instrumentalist account of epistemic reasons that, if successful, would refute “Instrumentalists Have too Few Reasons.” His strategy is inspired by two thoughts: that believing the truth is often very helpful, regardless of what your specific ends are, and that the concept of promotion is compatible with a very weak understanding of the promotion-relation.

To see the intuition behind his account, consider Mary, who desires to purchase a new pair of shoes. It’s not necessary that Mary has any true beliefs to achieve her goal: “she could happily succeed at buying a new pair of shoes [...], all the while believing that she is in the process of turning cartwheels down Broadway” (Schroeder, p. 114, 2007). Nonetheless, there are things that it would be useful for her to know: who sells shoes, what size she wears, and so on. Believing truly about these facts promotes her goal of buying a new pair of shoe. Schroeder thinks that this insight can be extended to account for all the epistemic reasons we take ourselves to have.

This first step of the account follows from a general analysis of reasons:

**Reason**  $r$  is a reason for  $a$  to  $\phi$  iff there is some  $p$  such that  $a$  has a desire whose object is  $p$ , and the truth of  $r$  is part of what explains why  $a$ ’s  $\phi$ ’ing promotes  $p$ .

Let’s grant this. To demonstrate that any arbitrary agent with any arbitrary goal has a reason to avoid being in error about any proposition, we need a theory about the promotion-relation. The theory of the promotion relation that Schroeder gives

is as follows:

**Promotion**  $a$ 's  $\phi$ -ing promotes  $p$  just in case  $Pr(p \mid a \phi\text{'s}) > Pr(p \mid a \text{ does nothing})$ .

It's worth noting that this means genuine reasons come cheaply. For example, suppose I'd like to see the Oscar Meyer Weinermobile. It's *very* unlikely that I will see it if I run out of my office right now. But I am more likely to see it if I run outside right now than if I do nothing at all. So—on this account—I have *a* reason to run out onto the street, one that is grounded in my desire to see the Oscar Meyer Weinermobile.

We can now see how the theory is supposed to account for epistemic reasons. The idea is that for *any* proposition, each of us has *some* aim that will be promoted if we believe in accord with the evidence with respect to that proposition. The too few reasons problem is solved by a theory that makes it *very* easy to have a reason.

Schroeder's account of epistemic reasons has been rejected by some for the fact that on his view we have *too many* reasons (e.g., Côté-Bouchard, 2015). If we account for the too few reasons problem by letting virtually anything count as a reason for anything, the complaint goes, this robs the instrumentalist of her ability to account for the distinctive nature of evidential reasons. If Schroeder's account entails that evidence for  $p$  is always a reason to believe it, but also that—say—counter-evidence that  $p$  is always a reason to believe  $p$ , then evidence loses its special status. And if this is so, the cure turns out to be no better than the disease.

I don't think, however, that this is the most promising argument against Schroeder's view. To contain the proliferation of reasons, Schroeder introduces a mechanism for weighing reasons. What is important is that regardless of the number of reasons we have, we need an account of how to properly weigh those reasons in deliberating about what to believe. If it turns out that the instrumentalist can give an account of

weighing that tells us that counter-evidentialist reasons carry no weight in doxastic deliberation, and that the evidentialist considerations do, then she can account for the distinctive status of evidential reasons.

A much larger problem for Schroeder's account is that despite its proliferation of reasons, it *still* has too few reasons. To see this, all we have to do is imagine the type of case that Kelly has us imagine: a case where an agent has evidence that supports taking some attitude towards some proposition, but the agent's taking that attitude would not promote *any* of the agent's desires. Cases like these will surely be less common on Schroeder's account, but they won't be *eliminated*. And so, the too few reasons problem is mitigated but persists nonetheless.

#### 2.2.2.2 Rule Instrumentalism, Modalized

Kornblith (1998) and Leite (2007) provide accounts of epistemic normativity that go in for rule instrumentalism and aspects of hypotheticalism. The view, let's call it *rule pragmatism*, holds that truth is uniquely and nearly universally helpful, but concedes that there are some instances where respecting the evidence won't us get what we want, and so takes a rule-based approach.

My main criticism of this view is derivative of what I said about rule instrumentalism. I have argued that if you go in for rule instrumentalism on account of exception cases, then you can only provide reasons to believe in those cases at a very high cost.

We can say a bit more about this view. Whether abiding by evidentialist norms is a good idea for an agent will, on this view, depend on contingent matters. This will spawn a rule-content problem: for some agents the belief-forming rules they should adopt may not be evidentialist in spirit. To build a simple case, consider an agent who only values her own subjective happiness. For this agent, the rule

**Evidentialism** always believe in accord with the evidence

will be inferior to

**Hedonism** if it's of obvious practical importance, believe on the evidence; otherwise believe whatever makes you most happy.

So, even if we grant that instrumentalists can help themselves to rule-based reasons, rule pragmatism cannot guarantee "Evidence Always Provides a Reason."

### 2.2.2.3 General Remarks

While it's true that in many instances believing in accord with the evidence will promote one's ends, it just isn't plausible that for every agent respecting the evidence—on every occasion or as a general rule—will promote their ends, whatever those ends are. Even if we stretch our understanding of promotion so that it is very inclusive, the problem persists. And the only other attempt to bridge that gap—the rule-based attempt—fails either because it cannot tell a plausible story about motivation or because the contents of its rules will not be thoroughgoingly evidentialist.

This concludes my argument against instrumentalism's hopes of rejecting "Instrumentalists Have too Few Reasons." At this point, the instrumentalist has two options: deny the seemingly unassailable "Evidence Always Provides a Reason" or admit defeat. In the next section I will demonstrate that the former option can be made plausible.

Before turning to my account it will be instructive to discuss an instrumentalist account of epistemic reasons that is sympathetic to the main idea of the paper so far. Sharadin (*forthcoming*) gives an account of epistemic reasons that ends up in the same position as the hypotheticalist, able to minimize the number of exception cases

we will encounter but unable to eliminate them. Seeing that the instrumentalist cannot eliminate exception cases, Sharadin endorses a “non-vindicating” account of epistemic normativity, an account that denies “Evidence Always Provides a Reason” and gives a debunking story about why we (incorrectly) believed it to be true.

The debunking story is very simple. On Sharadin’s account, exception cases are *extremely* uncommon. This is because he endorses a view very much like the hypotheticalist position considered above: On his account, for the vast majority of facts, believing those facts will promote *some* goal of ours. Given that it’s *overwhelmingly* often the case that believing in accord with the evidence will promote our ends, Sharadin argues, we simply overgeneralize and (incorrectly) think that it *always* provides a reason.

There are two issues with this story. I agree with Sharadin that there are plausible ways to deny “Evidence Always Provides a Reason”, but—somewhat paradoxically—I don’t think that it can be said that there can be cases where there is evidence that  $p$  and *no* reason to believe  $p$ . It is very difficult to state a non-question-begging reason to think this. However, the intuition in Kelly’s cases seems clear. Further, to say “I have evidence that it’s raining but no reason to believe it” borders on incoherence. It seems that the best interpretation of anyone who says this is that they misunderstand the nature or concept of evidence. Perhaps this is not the kind of consideration that would change the hearts or minds of those who accept the debunking story. But for those of us that don’t, it helps to identify a constraint on non-vindicating accounts: a non-vindicating account is plausible only if it can preserve the thought that wherever there is evidence that  $p$ , there is a reason to believe  $p$ .

The second issue presupposes the the first. Not only does it seem true that



there are reasons wherever there is evidence, but this seems to be non-contingently the case. That is, it seems that it is part of the very nature or concept of evidence that it provides a reason for belief. The intuition that drives “Evidence Always Provides a Reason” isn’t that evidence always provides reason *in actual cases*, but that it provides a reason *in all cases*.

The task, then, is to show how instrumentalists can show that necessarily, where there is evidence that  $p$  there is a reason to believe  $p$ . In the next section I show how this can be done.

## 2.3 The Compelling Argument, Revisited

At the beginning of this paper, I singled out genuine reasons as of interest, given our topic. I have not said much about the status evidence has when it does not provide a genuine reason. Obviously, it has the status of “non-genuine reason,” but this is too inclusive a label to be helpful, as it could be understood to include motivational reasons, causal reasons, explanatory reasons, and so on. We need something more informative than that.

Here is how I would like to set things up. Let’s call reasons that embody norms *normative reasons*—this is the class to which genuine reasons, legal reasons, and evidential reasons all belong. It excludes causal and explanatory reasons, like “reason” in “the reason my phone is broken is that I dropped in the toilet”. It’s worth noting here that my label “normative reasons” is fairly capacious. Commonly, “normative reason” is taken as a synonym for what I have called genuine reasons, which I’ll now refer to as “genuinely normative reasons.”

Let’s divide the class of normative reasons in two: in one subclass we have the reasons that embody standards one is robustly criticizable for failing to meet—the

now familiar *genuinely normative reasons*—and in the other we’ll have reasons that do not embody standards that one is rightly criticizable for failing to meet, the *merely normative reasons*.

I am now in a position to formulate a second possible interpretation of the The Compelling Argument:

### The Normative Compelling Argument

*Instrumentalists Have Too Few Normative Reasons* If epistemic instrumentalism is true, then it’s not the case that if an agent has evidence that Bertrand Russell is left-handed, she thereby has a **normative** reason to believe that he was left-handed.

*Evidence Always Provides a Normative Reason* If an agent has evidence that Bertrand Russell is left-handed, she thereby has a **normative** reason to believe that he was left-handed.

So, epistemic instrumentalism is false.

For ease of comparison, here is the version of the argument that we have been working with in the paper until this point (with updated terminology):

### The Genuine Compelling Argument

*Instrumentalists Have Too Few Genuine Reasons* If epistemic instrumentalism is true, then it’s not the case that if an agent has evidence that Bertrand Russell is left-handed, she thereby has a **genuinely normative** reason to believe that he was left-handed.

*Evidence Always Provides a Genuine Reason* If an agent has evidence that Bertrand Russell is left-handed, she thereby has a **genuinely normative** reason to believe that he was left-handed.

So, epistemic instrumentalism is false.

In discussions related to The Compelling Argument, it is not always clear which argument is at work. And, once we make this division it’s unclear whether either is sound.

As we have seen, the overwhelmingly popular strategy among instrumentalists has been to reject “Instrumentalists Have too Few Genuinely Normative Reasons” (which, if successful, would entail the falsity of “Instrumentalists Have too Few Normative Reasons”) by showing that there’s always a genuinely normative, end-given reason to respect the evidence. I have argued that this strategy is vexed: one thing that the Compelling Argument gets right is that it is implausible that every agent always has genuinely normative reason end-given reason to believe in accordance with her evidence.

For this reason, instrumentalists must reject “Evidence Always Provides a Genuinely Normative Reason.” If you have an intuition to the contrary, note that I am not advocating that instrumentalists reject “Evidence Always Provides a Normative Reason,” as any view that rejects this just gets the concept or nature of evidence wrong, and is false in virtue of this. Note also that accepting “Evidence Always Provides a Normative Reason” poses no problem for the instrumentalist, so long as she denies “Instrumentalists Have too Few Normative Reasons,” which—as I will argue for in a moment—she can do.

In what follows I will offer some reasons that speak against “Evidence Always Provides a Genuine Reason” but not “Evidence Always Provides a Normative Reason.” I think “Evidence Always Provides a Normative Reason” is right, and I think that a plausible account of epistemic normativity must account for this. In what follows, I show that there are at least two different (and compatible) accounts that instrumentalists *can* give for the premise. In giving these accounts, I will show that “Instrumentalists Have Too Few Normative Reasons” is false. So, by the end of the paper, I will have shown both The Genuine Compelling Argument and The Normative Compelling Argument to be unsound.

### 2.3.1 Evidence Always Provides a Genuine Reason?

There are two reasons to resist “Evidence Always Provides a Genuine Reason” that I would like to discuss, one that is internal to instrumentalism and one that is not.

The internal reason is that *if* we are confident in internalism, the first premise of Kelly’s *modus tollens* just is grist for the following *modus ponens*:

1. If epistemic instrumentalism is true, then it’s not the case that if an agent has evidence that Bertrand Russell is left-handed, she thereby has a genuine reason to believe that he was left-handed.
2. Epistemic instrumentalism is true.
3. So, it’s not the case that if an agent has evidence that Bertrand Russell is left-handed, she thereby has a genuine reason to believe that he was left-handed.

This, of course, can’t be a serious response to *Kelly*. But for those who find instrumentalism attractive, it is worth considering. I began this paper with a list of attractive features of instrumentalism. Most notably, instrumentalism gives a fully natural account of normativity and has an attractive story of motivation. Whether the above *modus ponens* speaks to you will depend on how deep your commitments to those ideas are and whether you think there are alternatives to instrumentalism that have these attractive features.

I now turn to a rejection of “Evidence Always Provides a Genuine Reason” that has wider appeal. It’s typically thought that there is a tight connection between what can be genuinely required of me and what I am able to voluntarily do; in a slogan, “ought implies can.”<sup>7</sup> There are many articulations of this intuitive idea. For now I’ll work with the following formulation:<sup>8</sup>

<sup>7</sup>For a defense of ought implies can, see Zimmerman (1996), Haji (2002), Streumer (2003), Vranas (2007), and Littlejohn (2012).

<sup>8</sup>The following argument has its origin in Alston (1989); for an extended commentary and defense see Côté-Bouchard (*forthcoming*).

**OIC** If my  $\phi$ 'ing is involuntary, then I can't be required to (refrain from)  $\phi$ 'ing.

It's also commonly held that

**INVOLUNTARINESS** We do not have voluntary control over our doxastic attitudes.

INVOLUNTARINESS needs little defense,<sup>9</sup> but an illustrative example might help explain why it's commonly accepted: if I stub my toe I can't but believe that I have stubbed my toe, that I am in pain, that I stubbed it on the bed post, and so on. There are, of course, debates over what the proper formulation of OIC is, whether any articulation of it is true, whether the proper formulation of the intuition behind it makes reference to involuntariness, and what involuntariness amounts to (I have left it undefined). Similarly, we can ask questions about INVOLUNTARINESS—do we have *some* degree of voluntary control over our doxastic attitudes? Do we lack control of *all* our doxastic attitudes? And so on. I will not wade into these debates here. I will simply note that it's generally thought that a principle in the vein of OIC is true, that a claim like INVOLUNTARINESS is true, and that this rules out the possibility of genuine normativity in the theoretical domain. This provides independent motivation for the instrumentalist's denial of "Evidence Always Provides a Genuine Reason".

### 2.3.2 Too Few Normative Reasons?

We began this paper with the clear intuition that evidence that  $p$  always provides a reason to believe  $p$ . An account of epistemic normativity should be able to account

---

<sup>9</sup>See Levy (2007) for a defense.

for this. In what follows, I provide two kinds of accounts that the instrumentalist can give.

I have said repeatedly throughout this paper that it's part of the *nature or concept* of evidence that it provides a reason to believe. The reason for the disjunction is that there is debate as to whether it's *metaphysically* necessary that evidence provides a reason or whether it's *conceptually* necessary that evidence provides a reason. Proponents of the former view are sometimes called *Teleologists*.<sup>10</sup> Proponents of the latter are sometimes referred to as *Normativists*. Building off of the accounts considered earlier in the paper, I'll show that instrumentalism is compatible with teleological and normativist defenses of "Evidence Always Provides a Normative Reason."

Before turning to the accounts, it's worth pausing to ask why instrumentalists need to give an account at all. After all, "Instrumentalists Have Too Few Normative Reasons" may be false simply because epistemic reasons have nothing to do with normative instrumentalism. The normative instrumentalist could, after all, maintain that she is only giving an account of the *genuinely normative* reasons. Anything else, she might say, is outside the realm of her concern.

Something about this seems unsatisfying. Normative instrumentalism is a theory of normativity, and so it seems fair to ask of the theory not just that it provide a theory of the reasons that are genuine, but what unites normative reasons. And if the normative instrumentalist can tell us why the agent in Kelly's story has

---

<sup>10</sup>To forestall any confusion, I would like to point out that there is another position about epistemic norms that is also sometimes called teleology. The position—call it *epistemic consequentialism*—is about the content of epistemic norms, and it parallels *ethical* consequentialism in structure. Where the *ethical* consequentialist thinks that proper conduct is about promoting an end—like overall utility—this view thinks proper belief is all about promoting an end, like that of maximizing the ratio of true to false beliefs you have. The main difference between teleology and epistemic consequentialism is that teleology says that evidential reasons are grounded in goals that are actually held by agents or subpersonal processes; epistemic consequentialism, on the other hand, gives us an account of epistemic value and what norms follow *if* we pursue that value.

a normative reason to believe in accord with the evidence, this would be one step towards rising to this challenge. Further, if the instrumentalist can give a compelling response, then any challenge posed by The Normative Compelling Argument can be diffused.

### 2.3.2.1 The Teleological Solution

One way to capture the idea that evidence necessarily provides a normative reason is to make the metaphysical claim that whenever an agent has evidence she has a normative reason to believe, and this is due to the nature of belief. Teleologists accomplish this by saying if an agent has evidence, she at least has a process that has the aim of regulating acceptances so that they are true.

Recall Velleman's proposal from above, which can be summarized with two claims: Belief is acceptance guided by the aim of thereby accepting a truth, and aims can be realized at the personal and subpersonal level. On this account, for any proposition that we have evidence for, there is a personal or subpersonal end we have that would be served by believing on the evidence. This, as we have seen, is not the sort of end that always gives us a genuinely normative reason. However, it can be used to vindicate and explain "Evidence Always Provides a Normative Reason" in instrumentalist spirit. The idea here is that normativity is all about promoting ends—that's what unifies normative reasons. What's genuinely normative *for an agent* is a function of that agent's desired ends. But this doesn't rule out the existence of *merely* normative reasons that spring from her subpersonal processes. So, instrumentalists can endorse a metaphysically necessary reading "Evidence Always Provides a Normative Reason" all while rejecting "Evidence Always Provides a Genuine Reason."<sup>11</sup>

---

<sup>11</sup>Velleman's account is thought by some—e.g., Shah (2003)—to suffer from a dilemma: it's either

### 2.3.2.2 The Normativist Solution

The above story, perhaps, is bit unsatisfying. We don't conceive of the claim that evidence provides a reason as a claim about mechanisms out in the world whose ends would be served by believing in accord with the evidence. Rather, the idea that evidence that  $p$  is a reason to believe  $p$  just is part of the *concept* of evidence. That is, if you don't think that evidence that  $p$  is—in some sense—a reason to believe  $p$ , you just don't have the concept of evidence. How might the instrumentalist account for this way of understanding “Evidence Always Provides a Normative Reason”?

The instrumentalist can account for this by understanding the concept of evidence as one that frames a certain kind of goal-oriented practice. Consider, for example, how William Alston characterizes the “epistemic point of view”:

Epistemic evaluation is undertaken from what we might call the “epistemic point of view.” That point of view is defined by the aim of maximizing truth and minimizing falsity in a large body of beliefs [...] [O]ur central cognitive aim is to amass a large body of beliefs with a favorable truth-falsity ratio. For a belief to be justified is for it [...] to be awarded high marks relative to that aim [...] [A]ny concept of epistemic justification is a concept of some condition that is desirable or commendable from the standpoint of the aim at maximizing truth and minimizing falsity (Alston, pp. 83-84, 1985).

What Alston is saying here is that we can make sense of epistemic justification from the perspective of a goal: that of having a large body of beliefs with a high truth-ratio.<sup>12</sup> While we might disagree with the specific goal that Alston has chosen

---

part of the nature of belief to be *very* sensitive to the evidence, or it is not. If it is, then certain mental states that count as beliefs will not—on Velleman's account—count as beliefs. Instances of wishful thinking, for example, will likely not count as beliefs. If, however, it isn't part of the nature of belief to be very sensitive to the evidence, then the teleologist is unable to explain *transparency* (the phenomenon that within the perspective of first-personal doxastic deliberation—deliberation about what to believe—the question *whether to believe that  $p$*  is transparent to the question *whether  $p$  is true*). I do not have the space to engage with this criticism here, but I do think that it can be responded to. For a response I am sympathetic to, see Steglich-Petersen (2006).

<sup>12</sup>Berker (2013a) and (2013b) argue that all the means-end accounts of epistemic norms, like the one given by Alston here, are false. I do not have the space to respond to Berker's arguments in



here, the general idea seems very plausible. Importantly for the instrumentalist position I am envisioning here, it's perfectly coherent to think of this goal as purely hypothetical, and the reasons the evidence gives as normative.

We may even be able to extend this idea to different normative domains. It seems plausible that normative domains—epistemology being a special case—are defined in terms of a goal. In epistemology, that goal is—perhaps—amassing a large body of beliefs with a favorable truth-ratio. In ethics, it might be maximizing overall utility or respecting persons. More generally, the instrumentalist could say, the normative reasons for some domain are the reasons that promote the goal that characterizes that domain, or that domain's "perspective." Those reasons are genuine when the end that characterizes the domain is adopted by the agent, otherwise they are merely normative.

This, I imagine, will be met with an objection. "Evidence Always Provides a Reason" is not a hypothetical statement. The intuition behind it was one that takes evidence to provide a reason, period. It is not the intuition that evidence provides a reason *if one has a certain goal*. This, however, does not strike me as problematic. To see why, consider an analogy. The rules of a game only apply if you have a goal, like winning at *chess*. But whether or not you have the goal of winning at chess, it is natural to speak as though chess-reasons are categorical. We say "don't castle in check" and not "don't castle in check *if you want to play chess*." This is because the rules of chess frame the practice of playing chess. Similarly, the normativist can say, talk of evidence frames the practice of epistemic evaluation—it simply presupposes the pursuit of the truth.

This might sound plausible, but is it an instance of instrumentalism? At the

---

this paper. However, I think that his arguments can be responded to. For a response that I am sympathetic to, see Ahlstrom-Vij and Dunn (2014).

beginning of this paper, I identified instrumentalism as the idea that your *genuine* reasons to depend on *your* ends. Normativism strays from this: it says your *normative* reasons are a function of *hypothetical* ends. While this view may not satisfy our definition of instrumentalism, it is compatible with instrumentalism and maintains instrumentalism's attractive features. It is perfectly natural, as it defines reasons in terms of natural entities. And, it respects the idea that our genuine reasons are reasons that spring from our ends.

### 2.3.3 Genuinely Normative (Non)evidential Reasons?

There is a question that I have yet to address: Is instrumentalism—epistemic or normative—compatible with the claim that evidential reasons are at least sometimes genuinely normative?

If instrumentalists can claim that reasons for belief are at least sometimes genuinely normative, this may bolster the view. We do seem to think that at least *some* beliefs are rightly criticizable. When I am making decisions that will greatly impact a loved one's well-being, for example, I am typically rightly criticizable if the beliefs that inform my decision do not reflect the evidence.

Fortunately, instrumentalists can allow for genuinely normative evidential reasons without contradicting our reasons for rejecting "Evidence Always Provides a Genuine Reason." I provided two reasons for rejecting this claim: one was a commitment to instrumentalism, the other a commitment to OIC and INVOLUNTARINESS. For proponents of the former, we could simply say that normative reasons to believe *p* are genuine when they are underwritten by our ends. For proponents of the latter, we would need to think that at least some of our beliefs are under our control (in the sense relevant for OIC). I do not have the space to explore all of the

possibilities here, but I see no reason to think that at least *some* of our beliefs are under our control. I may, for example, know that some complicated evidence I have speaks as to whether  $p$ , and I may be able to voluntarily decide whether or not wade through it. So it seems that I can choose between the attitude I currently have towards  $p$  and the attitude the evidence supports. Perhaps this is enough control to make genuine the reasons the evidence gives with respect to believing  $p$ .

If instrumentalists can claim that reasons for belief are at least sometimes genuinely normative, this may bolster the view. We do seem to think that at least *some* beliefs are rightly criticizable. When I am making decisions that will greatly impact anyone's well-being, for example, I am typically rightly criticizable if my beliefs do not reflect the evidence.

This raises another question: can we be rightly criticizable, per instrumentalism, for believing in accord with the evidence? It would seem so, and this may seem counterintuitive. We're always entitled to believe on the evidence, one might say. I'm not sure that this is right, however. It seems to be required of me that I think more highly of my friends and family than the evidence supports—that I give them the benefit of the doubt and their testimony more weight, for example. If I don't, I may be rightly criticized for being a bad friend or brother. It also seems required of me that I believe that my spouse and I will not be divorced, and that we will keep our vows. I should believe these things even though the evidence may not support it: the divorce rate for people like us hovers around 50%, and our vows were aspirational and idealistic.<sup>13</sup> And yet, it seems that it would be right to criticize me were I believe in accord with the evidence on these matters. Importantly, instrumentalism can explain these cases nicely. It's plausible that my life goes better when I trust my friends, my spouse, and the like.

---

<sup>13</sup>I owe this example to Maguire and Woods (*manuscript*).

It is important to remind the reader that this does not mean that we are always rightly criticizable when we believe on the evidence and against our ends. As I have stated, we often seem to lack control over what we believe. This, with OIC, can help to explain why in many cases we are not rightly criticizable for believing on the evidence. For example, if knowing the exact depth of the crevasse that I must cross would only be detrimental to my ends, I may not be criticizable if I believe on the evidence when confronted with a sign that marks its depth. This is because I can't but believe on the evidence when it confronts me in this way—in this instance believing on the evidence is more something that happens to me than something I do.

A perhaps surprising outcome of our discussion is that if instrumentalists allow some evidential reasons to be genuine reasons, then they must admit that we can be rightly criticized for believing in accord with the evidence. The instrumentalist could, of course, avoid this outcome by denying that we could ever have genuinely normative reasons for belief. But this would come at very high cost. In plenty of circumstances it's all too natural to hold agents accountable for believing on the evidence. This puts pressure on the instrumentalist to allow for genuinely normative evidential reasons. Luckily, the cost of allowing for them is not very high. This is because we do, in fact, require agents to believe against the evidence—we are, for example, to have confidence in our friends, family, and ourselves in ways that go beyond the evidence.

## 2.4 Conclusion

I hope that the arguments of section 2 lend credence to the claim that the standard instrumentalist approaches to accounting for "Evidence Always Provides a

Reason" are hopeless: it's just not plausible that agents have a special interest that could ground it, that believing truth is always helpful, or that a rule-based approach will can make either of these ideas work. I hope, however, that the argument of the previous section makes palatable the idea that evidence that  $p$  is not always a genuine reason to believe  $p$ . The only challenge that remains—in my opinion, at least—is to provide an explanation of evidential norms that is friendly to instrumentalism. I have shown that this can be done.

## Chapter 3

### Why Justify Belief?

**Abstract.** An unjustified belief is a flawed belief. Why is this? According to an intuitive and popular thought, the story—in sketch—goes as follows: Having true beliefs is the primary epistemic goal. Having justified beliefs is the best means to achieving that goal. So, having an unjustified belief is—from the epistemic point of view, at least—always a mistake. This is a promising beginning, but, as I argue, filling in the details is not as easy as it may seem. One popular path forward treats true belief the same way utilitarians treat utility: as something we should want more of. The other sees each belief as an attempt to get things right with respect to the proposition it's directed towards. Both face serious difficulties. The source of these difficulties, I argue, is that these accounts treat justification as a mere means to the primary epistemic goal. I propose an alternative account that treats true belief as the primary epistemic goal and treats justification as an end in itself. In virtue of this, the proposal captures the attractiveness of the intuitive thought we began with, while avoiding the difficulties associated with it.

#### 3.1 Introduction

There is an important connection between belief and some version of the following norm:

- (J) There is something wrong with unjustified beliefs.

Beliefs are judged against (J), they are also quite strongly regulated by it. Typically, there is something incompatible with judging that  $p$  while acknowledging in full

awareness that your belief  $p$  is not (epistemically) justified. This is so even when we take ourselves to have excellent practical reason to unjustifiably believe  $p$ . If, for example, believing  $p$  against the evidence is the only way for me to move on with my life after a traumatic experience, succeed at a long and difficult task, or win a large sum of money, I have excellent *practical* reason to believe  $p$  against the evidence. However, we seem unable to pull off this feat without engaging in self-manipulation. This makes (J) unlike many other norms. For suppose that

**(M)** There is something wrong with an action that violates someone's rights.

In typical cases, one is able to knowingly violate (M). If you think that (M) is false, plug in any plausible theory of right action you prefer and will get the same result. This highlights the uniqueness of (J)—it seems to have a special kind of regulatory control over belief. This cries out for an explanation.

The aim of this chapter is to make progress on two related questions. Call one of them *the source question*: What is (J)'s normative source? That is, what grounds (J)? Call the other *the motivational question*: why does (J) exert such a strong influence on belief? I argue that these two questions are connected in an important way, and that seeing this not only sheds light on (J), but the nature of belief.

This chapter proceeds as follows. In sections 3.3 and 3.4, I discuss two tempting first passes at answering the source and motivation question. The views, which I'll respectively call "epistemic consequentialism" and "epistemic teleology," both share in common a commitment to what I'll call means-ends and veritism.

*Means-Ends*: Epistemic norms arise from the instrumentally rational pursuit of some goal.

Instrumental rationality is the rationality one displays in adopting the means to one's ends.

*Veritism*: The goal whose pursuit gives rise to epistemic normativity is a strictly truth-oriented goal

Means-ends and Veritism both make intuitive sounding claims and enjoy wide support among epistemologists. While epistemic consequentialists and epistemic teleologists will differ in how they put these two intuitive ideas together, they share in common the same promising strategy:

- Step 1** Identify a goal that is constitutive and individuating of the epistemic point of view or belief.
- Step 2** Make a plausible claim about the requirements of instrumental rationality.
- Step 3** Make a substantive claim about the project of pursuing the ends mentioned in Step 1.
- Step 4** Derive (J).

I will argue that two tempting approaches to filling out the details in the above schema fail. These accounts fail as a result of making the satisfaction of (J) a *mere* means to achieving the proposed goal. Such views, I argue, are doomed to run afoul of (J). Defending this idea will take some time, but the basic thought is simple. If satisfying (J) is a mere means to some other end, there will be circumstances where violating (J) is a better means to that end than is observing (J).

I will then offer an alternative instrumentalist explanation of (J). The basic idea behind the account is simple as well. I will show that an account that holds that the goal that is constitutive and individuating of belief is one that builds a safety condition into the truth-goal itself. On this account, the aim of belief is have non-accidentally true belief.



## 3.2 Instrumentalism

Epistemic instrumentalism falls into a familiar metanormative tradition: call it *instrumentalism*, the view that normativity has its ultimate basis in the pursuit of (psychologically derived) ends. On some accounts, the operative ends are endemic to agency: part of what it is to be an agent, such accounts maintain, is to have certain ends. And that constitutive end of agency, these accounts maintain, has substantive consequences—that, for example, we ought to obey the categorical imperative. Other accounts are less optimistic that instrumentalism can vindicate the claim that all agents have moral reasons or obligations that can be derived from the constitutive goal of agency. Such views nonetheless maintain that normativity is to be understood on the means-end model (or something like it): what one ought or should do is a function of her ends, the empirical facts, the requirements of instrumental rationality, and nothing else.

Instrumentalists understand normativity as a perfectly natural phenomenon. A great number of philosophers take *naturalism*, the view that “all real properties are those that would figure ineliminably in perfected versions of the natural and social sciences,” as a fixed point in their theorizing (Shafer-Landau, 2003, p. 59). Insofar as one takes naturalism and normativity seriously, instrumentalism has serious attraction.

Instrumentalism is also well-poised to explain a curious feature of normativity: certain norms that apply to us have the power to motivate us. As much as I would like to keep as much of my money to myself as I possibly could, I can’t help but feel a very strong tug towards giving some of it to certain aid organizations. Why do I feel this tug to do what right? It’s because in addition to my commitment to my own happiness, there lie other commitments that I cannot shake: to treat like cases alike,

to intend to do what I believe I ought to do, and that there's no relevant difference between Singer's drowning child and the people I could assist by donating some of my money. Suppose, for the moment, that "treat like cases alike" and "intend to do what you believe you ought to do" are so basic and unassailable that they are constitutive of agency itself—that you, me, and everyone else are agents in part because of our commitment to these ideas. On this supposition, the fact that agents are motivated by Singer's argument for famine relief is no mystery at all: what we ought to do just is a function of the kind of things that typically motivate us—our commitments, desires, and so on.

I take it that these two features of instrumentalism make it attractive in general. But we can say more: it also seems to be a particularly nice fit for explaining epistemic norms. A great number of epistemologists think of (J) in instrumentalist fashion. To see why, one need look no further than Quine, who captured its spirit when he said that the

Naturalization of epistemology does not jettison the normative [...]. For me, normative epistemology is a branch of engineering. It is the technology of truth-seeking [...] it is a matter of efficacy for an ulterior end, truth.

To my ear—and hopefully to yours—something about this just sounds right. From the epistemic perspective, the pursuit of the truth is front and center. Furthermore, *if* one's pursuit is truth, it seems that (J) follows.

There is, of course, reason to temper any optimism we may have for the instrumentalization of epistemology. It is widely accepted that epistemic reasons have categorical force, that is, if one has evidence that *p*, one thereby has a reason to believe *p*. And one has this reason regardless of what her goals are. It's also widely accepted that there are propositions that some agents have *no* interest in believing truly. The conjunction of these observations is typically taken to entail that

any hopes of giving an instrumentalist account of epistemic normativity should be dashed. There is a lively debate on this topic, one about which I will have little to say in this chapter. The main reason for my silence is that the main interest of this paper lies elsewhere: the question at present is whether there is, in fact, a means-ends path from something like the Jamesian imperative—“Believe truth! Shun Error!”—to something like (J).

### 3.3 Epistemic Consequentialism

There is a prominent trend in epistemology—especially in formal epistemology—that treats truth or accuracy as a quantity that is to be promoted. These epistemologists think of true belief (or accurate credence) much in the same way that utilitarians think about well-being. In this section I will look at two manifestations of this idea. The first version of Epistemic Consequentialism that I will consider fashions itself after act utilitarianism. This simple view faces the same sorts of difficulties act utilitarians do; namely, it permits counterintuitive trade-offs. To reply to the challenge posed by these trade-offs, the epistemic consequentialist might attempt to develop a rule-based view. I argue that current approaches to a rule-based solution do not work.

#### 3.3.1 Direct Epistemic Consequentialism

In a passage representative of this position, William Alston states the basic thought nicely:

Epistemic evaluation is undertaken from what we might call the “epistemic point of view.” That point of view is defined by the aim at maximizing truth and minimizing falsity in a large body of beliefs (Alston (1985) *p.* 59).

Here, Alston expresses an intuitive thought. Filling in some details and expressed in terms of our schema, we get the following initially attractive explanation of (J):

**Step 1** The goal that is constitutive and individuating of the epistemic point of view is that of maximizing truth and minimizing falsity in a large body of beliefs.

**Step 2** Instrumental rationality requires: If your goal is to maximize the ratio of  $x$  to  $y$ , then  $\phi$  iff no other available option would bring about a greater net balance of  $x$  to  $y$ .

**Step 3** Of the available options, having (only) justified beliefs brings about the greatest balance of true to false beliefs.

**Step 4** So, (J).

While the articulations of steps 1-3 might vary from position to position, this serves as a nice representative for an intuitive explanation of (J).

The approach that the above is representative of—i.e., epistemic consequentialism—has received a lot of attention. This is on the one hand because of its fruitfulness (e.g., the accuracy-first approach to bayesian epistemology) and on the other because of its penchant for endorsing a certain kind of problematic trade-off that we will discuss shortly.

It's tempting now to give a general definition of consequentialism, so that we can understand what, if anything, epistemic and ethical consequentialism might have in common. I will not attempt to give such a definition here. The reason for this is that there's very little agreement about what it takes for a theory to count as an instance of consequentialism—ethical or otherwise:

Some maintain that a consequentialist is committed to understanding what is right or obligatory in terms of what will maximize value (Smart and Williams 1973, Pettit 2000, Portmore 2007). Still others maintain that a consequentialist is one who is committed to only agent-neutral,

rather than agent-relative prescriptions (where an example of an agent-relative prescription is one that instructs each person *S* to ensure that *S* not lie, whereas an agent-neutral prescription instructs each person *S* to minimize lying) (McNaughton and Rawling 1991). And finally, some maintain that what is distinctive about consequentialism is the lack of intrinsic constraints on action types (Nozick 1974, Nagel 1986, Kagan 1997). (Dunn, (2015))

This, however, does not make the classification “epistemic consequentialism” useless.<sup>1</sup>

The views that I identify as instances of epistemic consequentialism bear a very close resemblance to uncontroversially consequentialist views, such as act utilitarianism. That there is a close analogy between these views will suffice for our purposes. Where the act utilitarian thinks

*Act utilitarianism.* *S*'s  $\phi$ 'ing at *t* is (morally) right *iff* no other act available to *S* at *t* would bring about a greater net balance of pleasure over pain.

The epistemic consequentialist might hold:

*Direct epistemic consequentailism.* *S*'s believing *p* at *t* is justified *iff* no other belief available to *S* at *t* would bring about a greater net balance of true belief over false belief.

The problem with direct epistemic consequentialism is that it is susceptible to counterexamples. Consider:

*Scientist.* I am a scientist interested in getting a grant from a religious organization. Although I think that belief in the existence of God is manifestly irrational (from the epistemic perspective), I discover that this organization will give me the grant only if it concludes that I am religious. [...] I am such a terrible liar that unless I actually get myself to believe in the existence of God they will discover that I am an atheist (Fumerton, 2001).

The direct epistemic consequentialist would have it that I am epistemically justified in believing God exists. But I am clearly not epistemically justified in believing in God and because of this, direct epistemic consequentialism is false.

---

<sup>1</sup>I will have more to say about this label later in section 7.2.

### 3.3.2 Indirect Epistemic Consequentialism

There are, of course, moves that epistemic consequentialists can make in order to avoid getting the wrong result in Scientist. Before discussing those moves, let's diagnose what went wrong in Scientist. Direct epistemic consequentialism makes the overall epistemic consequences of a belief relevant to that the justificatory status of that belief. But this, as the counterexample teaches us, is not relevant to the justificatory status of a belief. Whether I'm epistemically justified in believing some proposition is not a function of whether that belief will contribute to my believing other true propositions. In a slogan, direct epistemic consequentialism does not respect the separateness of propositions (Berker (2013a)).

The problem of disrespecting the separateness of propositions closely mirrors the problem act utilitarians have with respecting the separateness of persons. In the case of utilitarianism, one strategy for avoiding this issue is to endorse an indirect utilitarianism, which judges actions in reference to the rule or system of rules that recommends that action, and judges rules in terms value-promotion. A promising strategy for the epistemic consequentialist, then, may be to mimic this strategy and develop an indirect theory of her own.

One such strategy, which has been explicitly endorsed by some reliabilists, is to say that the reliability of a process is to be determined on consequentialist grounds, but the status of any given belief as (un)justified solely depends on the status of the process that produced it. Returning to the analogy to utilitarianism, to maintain the separateness of persons, the rule utilitarian might say:

*Indirect utilitarianism.*  $S$ 's  $\phi$ 'ing at  $t$  is right iff the internalized set of rules that caused  $S$  to  $\phi$  at  $t$  is right,

where

Set of rules  $r$  is right *iff*  $r$  tends—when internalized by a suitable portion of the population in a suitable range of circumstances—to lead to a balance of pleasure over pain that is greater than some threshold,  $T$ .

Inspired by this, the epistemic consequentialist may endorse a view like the following:

*Indirect epistemic consequentialism.*  $S$ 's belief that  $p$  at time  $t$  is justified *iff* the belief-forming process that caused  $S$  to believe  $p$  at  $t$  is reliable,

where

Belief-forming process  $b$  is reliable *iff*  $b$ 's immediate outputs tend—when employed in a suitable range of circumstances—to yield a balance of true to false belief that is greater than some threshold,  $T$ .

These ideas can be incorporated in to our schema as follows:

- Step 1** The goal that defines the epistemic point of view is that of maximizing truth and minimizing falsity in a large body of beliefs.
- Step 2** Instrumental rationality requires:  
 If your goal is to maximize the ratio of  $x$  to  $y$ , then  $\phi$  *iff* your  $\phi$ 'ing is caused by an approved entity,  $e$ .  
 (An entity,  $e$ , is *approved* *iff* employing  $e$  tends—in circumstances  $c$ —to lead to a high ratio of  $x$  to  $y$ .)
- Step 3** A belief is caused by an approved entity *iff* it is justified.
- Step 4** So, (J).

Now that we have an indirect epistemic consequentialism that looks at the immediate outputs of processes to determine their reliability, we have a version of epistemic consequentialism that can get the right result in Scientist. For, say the scientist formed the belief as a result of wishful thinking—a paradigmatically unreliable process—then we would get the result that the belief is unjustified, regardless of whether the belief in God would lead to nice consequences.

It has been conjectured that even indirect epistemic consequentialism will have trade-off problems; we just need to be inventive about how to build the cases (Berker 2013a, b). Indirect epistemic consequentialism, for example, will have troubles with cases like this:

*Prime Numbers.* Suppose the following is true of me: whenever I contemplate whether a given natural number is prime, I form a belief that it is not. “Is 25 prime? No, it is not.” “Is 604 prime? No, it is not.” “Is 7 prime? No, it is not.” Let us also stipulate that this is the only cognitive process by which I form beliefs about the primeness of natural numbers. (I’m a simpleminded kind of guy.)

Because the ratio of prime to composite numbers approaches zero as  $n$  approaches infinity, the direct outputs of the this process will tend to be true and thus the process reliable. We’re meant to have the intuition that the beliefs formed on this basis, however, are unjustified and thus indirect epistemic consequentialism false.

The prime numbers case is clever, but I do not think that the intuition appealed to in this case can do the work requested of it. My concern is that the case relies essentially on on the stipulation that “this is the only cognitive process by which I form beliefs about the primeness of natural numbers.” Reliabilists have—to deal with defeaters—added conditions like the following to their account:

RULE.  $S$ ’s belief that  $p$  is justified *iff* it was produced by a reliable process, and there is no other reliable process available to  $S$  such that, had  $S$  used that process,  $S$  would not believe that  $p$  (viz. Goldman 1979).

This reliabilist theory can be understood in consequentialist terms: we could hold that a belief is justified only if it is approved by a reliable rule, RULE is reliable<sup>2</sup>, and a rule is reliable *iff* the doxastic attitudes it recommends tend to be true. Now reconsider prime numbers, but keep in mind that we drop the stipulation of simplemindedness. Our indirect epistemic consequentialism gets the right result in the

<sup>2</sup>It may sound like calling rules reliable commits some kind of category mistake—processes, and not rules—are the kinds of things that are (un)reliable, one might think. Goldman, however, has provided a framework for assessing the reliability of rules (viz. Goldman (1988)).



case. For the above reasons I will suppose that indirect epistemic consequentialism can be tweaked so that it is not subject to trade-off style counterexamples.

Even if we grant this, however, the proposal suffers from a serious problem. For suppose that we find ourselves in a Scientist-like case, where we are offered a trade-off: if we believe in God this will cause us to have a much higher ratio of true-to-false beliefs or will cause us to have a very high ratio of true-to false beliefs in a greater number of propositions. Indirect epistemic consequentialism will tell us that we shouldn't believe in God, which seems to be in tension with the attractive thought that motivates epistemic consequentialism, i.e., that the epistemic perspective is one that concerns itself with the goal of amassing a large number of mostly true beliefs. Note that this seems to put the indirect epistemic consequentialist in roughly the same awkward dialectical position that indirect utilitarians seem to find themselves in—having given up on a direct consequentialism, she seems now to be guilty of “rule worship”(Smart (1956)).

It's worth pausing for a moment to explore the parallel between the charge of rule-worship in the case of rule utilitarianism and in the case of indirect epistemic consequentialism. Rule utilitarianism and indirect epistemic consequentialism bear deep structural similarities. And both take it that instead of judging beliefs or actions in terms of their consequences, we ought to judge them against rules that would in general promote this end. This, of course, raises the threat of incoherence: if I obey rule *R* to promote good *G*, and I know that violating *R* will in this case maximize *G*, it's incoherent to obey *R* in this case.

The rule-worship objection to utilitarianism is an old one; given the fact that there are still rule utilitarians, we might think that there is a solution to the utilitarian rule-worship problems that the indirect epistemic consequentialist can help herself to. My sense is that this strategy cannot work. The reason for this is that there

are at least some cases where sacrificing one to save the many seems like the right thing to do. Consider, for example, a case where the only way to stop a train from hitting a button that will make a bomb go off in a crowded city is to flip a switch that will divert the train onto a track with exactly one person on it. Many of us, even those of us with strongly deontological intuitions, tend to think that in cases like this it is okay to kill the one to save the many. And this is just as it should be. The rule utilitarian has good reason to think that in the long run, general rules that do not require one to maximize utility are preferable to rules that are not. Yet, if she does not make space for some caveats—e.g., in cases where sacrificing the one to save the many is the only way to prevent disaster—then there seems to be some incoherence in her view. Luckily for the utilitarian, we are okay with *sometimes* making inter-personal tradeoffs.

Unfortunately for the indirect epistemic consequentialist, our intuitive notion of JUSTIFICATION seems never to permit inter-propositional trade-offs. And yet, if one is a consequentialist—rule or otherwise—to seem that if  $\phi$ 'ing will lead to extremely good results, then one should at least permit  $\phi$ 'ing. However, the downstream consequences of holding a belief never affect its status as epistemically (un)justified. For this reason, the epistemic consequentialist is pushed in two incompatible directions. She must, if she is to give an account of our intuitive notion of justification, hold that epistemic trade-offs are never epistemically justified. But this, from the point of view of her normative foundation—i.e., consequentialism—seems objectionably ad hoc. However, she must, per her consequentialism, permit such trade-offs. This, of course, means that she has given up on vindicating (J), or at least any intuitive version of (J).

We began with the idea that that the “epistemic point of view” is defined by the aim of maximizing truth and minimizing falsity in a large body of beliefs. We

can think of the versions of veritistic epistemic consequentialism as explications of possible norms that would follow were we to adopt this goal.

If we go in for the version of epistemic consequentialism that fits most naturally with the intuition that motivates it, either (J) is false or we end up with an incorrect theory of justification. If we understand justification as the naive consequentialist does, (J) is true but only because JUSTIFICATION has been (incorrectly) reconceptualized. If we have the commonsense notion of JUSTIFICATION at play, then (J) comes out false—there will be beliefs that are not we are not justified in believing that we should, per direct consequentialism, adopt. Either way, the direct approach seems hopeless.

If, on the other hand, we adopt a view like the indirect version of epistemic consequentialism described above, we end up with an account that can maintain (J) while maintaining an intuitive picture of justification. This, however, comes at a high cost. The view tells us that we ought, in virtue of our pursuit of truth, only believe the outputs of reliable processes or rules. But when we have an opportunity to promote our truth-goal that involves violating (J), we ought not violate (J). This, in effect, is to give up on the promising schema. So, on both versions of consequentialism considered here, we don't have an explanation of (J).

### 3.4 Epistemic Teleology

There is an instrumentalist approach that is quite separate from epistemic consequentialism that derives epistemic normativity from belief's "aim." It is worth noting that while accounts that focus on the aim of belief are diverse, all share a common commitment to the idea that a certain kind of "aiming" is constitutive and individuating of belief as a mental state. I have chosen to call these accounts

instances of “epistemic teleology” because all versions of this proposal appeal to the pursuit of a goal as the source of epistemic normativity. For now, we’ll focus on interpretations that understand the nature or concept of belief as aiming defined by the pursuit of a goal:

*Epistemic Teleology*: part of what it is to believe  $p$  is to believe  $p$  with the purpose of believing  $p$  only if  $p$  is true.

We are now in a position to appreciate how epistemic teleologists differ from epistemic consequentialists. Where a consequentialist’s verdicts depended—directly or indirectly—on the overall value of states of affairs, the epistemic teleologist looks only to whether *each* belief succeed by the lights of its own goal.<sup>3</sup> Importantly, this difference marks a major advantage that epistemic teleologists have over epistemic consequentialists. We noted in the last section that epistemic consequentialists have a problem with epistemic trade-offs. Teleologists have some hope in keeping propositions separate. If *each* belief is beholden to its own aim and nothing else, whether believing  $p$  will lead me to believe other truths is ancillary to whether I should believe  $p$ . Now we can ask whether we can get (J) from something like (T): <to believe  $p$  only if  $p$ >.

While there may be differences about how, exactly, the epistemic teleologist goes from her favored version of (T) to her favored version of (J), the basic arc of the story will be the same: when we believe that  $p$  we—in some sense—aim to do right by (T), and we can’t aim to achieve (T) without thereby aiming to do right by a norm like (J). This is because—roughly—(J) is the the proper means to doing right by (T). In terms of our schema:

---

<sup>3</sup>Clarification: we could understand one group as a subset of the other. It’s most common to think of consequentialists as a certain kind of teleologist. We can go along with tradition and speak this way so long as we keep in mind that in this paper “teleologist” means “non-consequentialist telologist”.

**Step 1** One's goal in believing  $p$  is to believe  $p$  only if  $p$ .

**Step 2** Instrumental rationality requires: If your goal is to believe  $p$  only if  $p$ , then believe  $p$  only if you are justified in believing  $p$ .

**Step 3** –

**Step 4** So, (J).

While seldom made explicit, I think that this captures story of how the teleologist moves from (T) to (J).

The above laces together means-ends and veritism in a way that is intuitive and avoids trade-offs. There is, however, a problem with this story. What we are looking for is an explanation of (J) that locates it in a familiar pattern, like <if your goal is to  $\phi$  and  $\psi$ 'ing is a necessary means to  $\phi$ 'ing, then  $\psi$ >. The stipulation at step two—i.e., If your goal is to believe  $p$  only if  $p$ , then believe  $p$  only if you are justified in believing  $p$ —does not give us this. Rather, it brutally asserts the connection that we are looking for.

### 3.4.1 Completing Step 2

So, what might we use at step two? In what follows I will consider three tempting options and then argue against them. Following that, I will then, in 3.4.2, give an argument to the effect that step two *cannot* be filled in such that we get (J) at step four.

#### 3.4.1.1 Requirement N?

To begin our exploration of principles that could do the work required at Step 2, we might turn to an uncontroversial principle of instrumental rationality, like

**N:** If your goal is to  $\phi$  and  $\psi$ 'ing is a necessary means to  $\phi$ 'ing, then  $\psi$ .

The problem with this proposal should be fairly obvious. On any commonsense understanding of the concept, it's possible to have an unjustified true belief; I might believe against the evidence but luck my way into believing truly. And, for this reason, **N** can't do the work required of it here.

#### 3.4.1.2 Requirement B?

While justification isn't a necessary means for truth, we might say that we do *believe* that having a justified belief towards  $p$  is the most reliable means for achieving true belief. In other words,

**B:** If your goal is to  $\phi$  and *you believe*  $\psi$ 'ing is a necessary means to  $\phi$ 'ing, then  $\psi$ .

The problem with this proposal, however, is that many agent have irrational beliefs about justification. For example, consider the counterinductivist, who thinks that if  $A$ 's have been followed by  $B$ 's in the past, we should conclude that the next  $A$  will *not* be followed by a  $B$ . Unless we are ready to say that such agent's beliefs are justified when they reason by counterinduction, we should not endorse this solution. Further, I do not think we are ready to say that counterinductivists are justified when they deploy counterinduction; so, I do not think we should endorse **B**.

#### 3.4.1.3 Requirement R?

Seeing that **B**'s pitfall was its subjectiveness, let us try a solution that is stronger than **N** but that relies on an objective standard. The following principle meets such a description:

**R:** If your goal is to  $\phi$  and you *rationally believe*  $\psi$ 'ing is a necessary means to  $\phi$ 'ing, then  $\psi$ .

To know whether **R** can save the day, we need to know what rationality amounts to in this context. Let us explore two hypotheses.

Let's first explore the idea that "rationally believe" means something minimal, such as "to believe without contradiction that". The problem with this proposal is that it is not a substantial improvement on *B*. The counterinductivist can believe without contradiction with that counterinductivism is the best way to form predictions.

Let us now suppose that "rationally believe" means something more substantive, such as "believe in accord with the evidence that". This is clearly problematic because, like the proposal we began with, it simply assumes what is meant to be shown, i.e., (J).

### 3.4.2 The Aim of Guessing

The process of proposing principles and generating arguments against them could go on indefinitely. In this section I aim to circumvent such a discussion by proposing an argument to the effect that *no* principle of instrumental rationality can fill in step two in a satisfactory way.

Belief is not the only attitude whose aim is (T). Consider guessing. It's plausible that the aim of guessing is truth; after all, if you guess falsely, there is something wrong with your guess (Owens (2003)). In other words

(**T<sub>g</sub>**) One's goal in guessing that  $p$  is to guess that  $p$  only if  $p$ .

If (J) is true in virtue of belief's aiming at truth, then the following is true in virtue of (**T<sub>g</sub>**)

(J<sub>g</sub>) There is something wrong with unjustified guesses.

This, however, is problematic, and because (J<sub>g</sub>) is false.

Even though guessing aims at the truth, it's false that a guess that is not epistemically justified is thereby flawed. To see this consider two different scenarios. In the first, you are debating with a friend about who will win the next election. You think that there is not enough evidence to decide the case—the most sophisticated analyses agree that it's a toss-up. Your friend, however, believes—as a result of wishful thinking—that her candidate will win. When pressed on this, she cites what she takes to be evidence: that a certain respected outlet's models favor her candidate by a narrow margin, and so on. Even if your friend's candidate wins, it's clear that she is still subject to criticism. Her belief is flawed regardless of whether her candidate wins.

Imagine, now, a separate scenario. Everything is the same as before, except you both recognize that no one was in a position to know which candidate would win. You decide nevertheless to guess for fun. Your friend guesses that her candidate will win, and she guesses this way just because it's her favored candidate. As before, her candidate wins. It's clear that in this second case that there is nothing wrong with her guess. This is explained by the fact that (J) is true and (J<sub>g</sub>) false.

While I think that the above argument works, there is a complication that I have left out that I now must mention. A guess is most naturally conceived of as an act, where a belief is more naturally conceived of as a state. Some philosophers, e.g. Velleman and Shah (2005), point out this disanalogy as a means for evading the above objection. This response, however, only moves the bump beneath the carpet, so to speak. For in deliberating about whether to believe that *p*, (J) frames my deliberation. And in this mode, *judging* (an act) that *p* constitutes a mistake. The



best explanation of this mistake is (J). Yet, in the mode of first personal deliberation, the fact that (T) frames our deliberation can explain (J) only if we use some premise like the principle we assumed in our first pass at step two. But as we have just seen, this cannot be made to work.

Let us briefly turn to the motivational question. Epistemic teleologists, unlike consequentialists, concern themselves with the motivational question. While stories differ from person to person, the heart of the intuition behind these views can be brought out by considering a very simple account of belief's aim. Suppose that we thought that in deliberating about whether  $p$  we take, as our goal,  $\langle$ to believe  $p$  only if  $p$   $\rangle$ . This, intuitively, saddles us with something like (J). Given that we have excellent reason—from our own point of view—to obey (J), this explains strong influence (J) has over our deliberation.

This story sounds intuitive enough, but the above reflections should give us pause. If having (T) as one's goal explains (J)'s influence over belief, then having  $(T_g)$  as one's goal should bring it about that  $(J_g)$  has a similar influence over guessing. But whatever influence  $(J_g)$  has over guessing, it is quite different from the influence (J) has over belief. To see this, just return to the election case. I am rationally confident that the race is too close to call. I cannot, while in full appreciation of my evidence, believe that the democrat will win. Why is this? It seems that it's because (J)—somehow—steps in and prevents me from doing so. But now let's contrast this with a different case. Again, I am rationally confident that the race is too close to call. When prompted, I nonetheless *guess* that the democrat will win. Even if I am committed to  $(T_g)$ ,  $(J_g)$  fails to get the same kind of grip on me that (J) has. This, I take it, shows that epistemic teleologists have failed to answer the motivational question.

### 3.5 What Went Wrong?

I'd like to pivot now to make a general diagnosis as to why neither of the above accounts seems to work. As I have noted, both approaches that I have considered work with a means-end model to explain the significance of justification. That is, both specify some desirable end—one that may be a good candidate as constitutive of the “epistemic perspective” or the aim of belief—and purport to show that if one takes that end as her goal, then—via some line of means-end reasoning—it is also proper form justified beliefs. In the case of epistemic consequentialism, we saw that to get from the goal of having a high ratio of true to false beliefs to (J), we needed to add something to the picture, something that made it the case that the proper way to evaluate belief-forming methods is in terms of their direct outputs. What needed to be added could be conceived of one of two ways. We could conceive of the missing ingredient as enriching the goal itself: the goal wasn't *just* to have a high truth-ratio, but that it was a high truth-ratio gotten in a certain way (i.e., via reliable processes). But this is a far cry from the thought that the value of true belief is prior to the value of justification. We could alternatively conceive of the added ingredient as enriching the means-end model: if  $x$  is your goal, then the proper way to promote  $x$  to via the selection of rules in a very particular way. This, however, strains the intuition behind the mean-end model and is ruled out by the argument of 3.4.2.

The epistemic teleologist faces a similar issue. We saw that something was wrong with the derivation of (J) from (T). To get to (J) we needed to add a premise that was too particularistic for our ends. Further, given the parallel between the derivation of (J<sub>g</sub>) from (T<sub>g</sub>), it seems like the best option for the teleologist can't be to enrich the means-end model. This is because unless she modifies (T), the parallel

principle at step two will either over generalize (and allow us to derive  $(J_g)$  from  $(T_g)$ ), else it will not be powerful enough to get us from  $(T)$  from  $(J)$ .

Despite all of this, I find features of the views considered compelling, and I do not yet want to give up on the idea that belief, in some sense, has truth as its goal. The above reflections show us a promising way forward: to preserve the intuitiveness of veritism and means-ends, we might enrich the truth-goal in a way that maintains the intuitive idea that belief is, in some sense, directed at the truth, in so doing we may have a goal that is rich enough that, via means-end rationality, actually entails  $(J)$ .

### 3.6 Safely Pursuing the Truth

As the election example reveals, one difference between guessing and believing is the amount of risk the norms involved permit you to take on. A guess that lacks justification can nonetheless turn out to be fully correct. Not so, it seems, for belief. Let us suppose that this is because the constitutive aim of belief isn't  $(T)$ , but

**(T\*)** <to believe  $p$  only if  $p$ , and not easily would  $Bp$  have been false.>

A few notes about this formulation are in order.

First, “not easily would  $Bp$  have been false” is in need of interpretation. To my mind, there are two sensible ways to understand this condition. I'll briefly discuss each in turn.

We might understand “not easily would  $Bp$  have been false” subjectively. We may, for example, understand “not easily would  $Bp$  have been false” as true when  $Bp$  is *unsurprisingly* true relative to a certain perspective. A plausible candidate for this perspective is one defined by a set of norms regarding the proper maintenance

of our beliefs. These norms may be those of the epistemic community we belong to, or the ones we have internalized as part of our education and upbringing.

Alternatively, we might understand “not easily would  $Bp$  have been false” objectively. That is, we might understand “not easily would  $Bp$  have been false” as true when  $\langle (Bp \ \& \ p) \ \& \ (Bp \ \Box \rightarrow \ p \ \& \ Bp' \ \Box \rightarrow \ p') \rangle$ , where, roughly,  $p$  and  $p'$  are close variations of one another (e.g., one bird flies past my head vs. two), and  $Bp$  and  $Bp'$  are generated by the same belief-forming processes (e.g., perception).

On some interpretations,  $(\phi \ \Box \rightarrow \ \psi)$  is true at  $W$  iff (a) some world in which  $(\phi \ \& \ \psi)$  is true is closer than any world where  $(\phi \ \& \ \neg\psi)$ , and (b) the closest world to  $W$  is  $W$ . This interpretation will not work for “not easily would  $Bp$  have been false.” (a) and (b) are true *any* time you believe truly and so for these purposes are satisfied too easily. Were we to accept this reading, if you believe truly but contrary to the evidence, you will satisfy (T\*). For these reasons, I favor an interpretation where  $(\phi \ \Box \rightarrow \ \psi)$  is true iff  $(\phi \ \& \ \psi)$  is true in the actual world,  $W$ , and all worlds “close” to  $W$ .

We still need to determine what makes a world “close.” I will not venture a theory here. I can, however, point to factors that are relevant—the belief in question ( $Bp$ ), the time ( $t$ ), the agent who holds  $Bp$  ( $A$ ), and the method by which the agent formed the belief ( $M$ ) (Rabinowitz, 2017). In the end, though, we’ll rely on intuition to tell whether one world counts as close to another.

Second, I hope that “not easily would  $Bp$  have been false” does not become a distraction. “Not easily would  $Bp$  have been false” and my interpretation of it are mainly in place to help fix ideas. The idea is to have something like (T\*) on the table in order to make good on the main claims of this paper: we can give a plausible instrumentalist account of (J) without running into the issues associated with epistemic consequentialism or the version of epistemic teleology

just considered. With this noted, I will treat (T\*) as a hypothesis. I will show that it has promising features that make it a serious improvement on the accounts we have considered so far.

Let us begin by discussing how our new account can explain (J). On the above conception, we aim not only to believe truly, but in a way that is safely true relative to the nearby possibilities. Beliefs that are the product of belief forming methods that canonically confer justification—e.g., perception, memory, and so on—are safe in just this way. Moreover, having justified beliefs is a necessary condition on having beliefs that are safe from this kind of error. In other words:

**Step 1** One's goal in believing  $p$  is <to believe  $p$  only if  $p$  and not easily would  $Bp$  have been false>.

**Step 2** Instrumental rationality requires: If your goal is to  $\phi$  and  $\psi$ 'ing is a necessary means to  $\phi$ 'ing, then  $\psi$  (i.e., requirement **N**).

**Step 3** Satisfying (J) is a necessary means to satisfying (T)\*.

**Step 4** So, (J).

It is my hope that I have made the four steps plausible. I do not take it that I have fully explained each step. However, I do think that I have said enough to motivate them. Further, I take it that the remaining sections of the paper fully illuminate any obscurities that I have not yet addressed.

I now briefly turn to the motivation question—why is it that, typically, we cannot in full awareness believe against the evidence? Above we considered an account according to which it's part of the nature of belief to be sensitive to (T). As the guessing example shows, this does not suffice to answer the motivation question. But now let's replace the hypothesis that it's part of the nature of belief to be sensitive to (T) with the hypothesis that it's part of belief's nature to be sensitive

to (T\*). Given that satisfying (J) is a necessary means to satisfying (T\*), our new hypothesis predicts that beliefs will display sensitivity to (J).

## 3.7 Applications

I'd like to briefly consider a few applications of the safety account of the aim of belief. My reason for considering applications is two fold: considering them helps us better understand the proposal, and its applications provide reasons to accept it.

### 3.7.1 Evidentialism

According to a compelling account of justification,

EVIDENTIALISM *S's belief that  $p$  at time  $t$  is justified iff  $S$ 's evidence as  $t$  supports  $p$ , and  $S$  believes  $p$  on the basis of evidence that supports  $p$ .*

EVIDENTIALISM, one might think, cannot be captured using the safety account. After all, it's quite easy to come up with cases where we believe on the evidence and yet our belief could easily have been false. Further, we can imagine cases that meet this description where believing against the evidence would seem like the correct option, per (T\*). To see this, one need look no further than to the familiar case of Norman the Clairvoyant (Bonjour (1980)):

Norman has a reliable clairvoyance faculty. However, he has evidence against the possibility of clairvoyant power and against his possessing a faculty of clairvoyance. One day his clairvoyance faculty produces in him a belief that the President is in New York City.

It would seem that on the above account, Norman's belief does not stand in need of correction. Further, it would seem to imply that Norman's bringing his belief in line with his evidence would constitute a mistake.

This may be true on the objective reading of “not easily would  $Bp$  have been false,” but it needn’t be true on the subjective reading. And, presumably one perspective from which it might make sense to evaluate Norman’s belief is that of an evidentialist standard.

### 3.7.2 Process Reliabilism

According to recent influential argument, process reliabilism is it is an instance of epistemic consequentialism and this gives us sufficient reason for rejecting it (Berker (2013b), 2013a). More specifically:

1. If process reliabilism is true, then epistemic consequentialism is true.
2. Epistemic consequentialism is false.
3. So, process reliabilism is false.

However, if process reliabilism is false, it isn’t false for these reasons.

I have given something of a case for premise 2., so the only premise that needs to be explained is premise 1. Earlier, I was reluctant to give a definition of “consequentialism.” My reason for this was that there was little consensus of what this term means. The above argument is taken from Berker (2013b), who understands having the following three features as sufficient for being an instance of consequentialism:

*A theory of final value*, which specifies certain states of affairs that have value as ends in themselves, and other states of affairs that have *disvalue* as *ends in themselves*.

*A theory of overall value*, which assigns a value ranking to any entity that *conduces toward or promotes* those states of affairs which, according to the theory of final value, have value or disvalue as ends in themselves, and

*A deontic theory*, which assigns deontic properties on the basis of the theory of overall value.

The issue that I have with the above is that it's not clear to me what it means for a theory to have this structure. With little contortion, it seems to me that one could argue that deontological theories, e.g. Kant's ethics, have this structure:

Theory of final value<sub>Kant</sub>: States of affairs where persons are respected have final value; states of affairs where persons are not respected have final disvalue; nothing else has final (dis)value.

Theory of overall value<sub>Kant</sub>: Maxim  $m$  has positive overall value if, when acted upon, it realizes a state of affairs where no persons are disrespected.

Deontic theory<sub>Kant</sub>: An action is right *iff* its maxim has positive overall value.

In any event, I will not proceed by way of establishing a definition of consequentialism. Rather, I will show that an account that uses (T\*) to explain the significance of reliability can avoid the problems that I associated with epistemic consequentialism. Whether the safety account ultimately is an instance of epistemic consequentialism is a question I will little to say about. For my purposes, the important point will be that the explanation of the significance of reliability that it can provide gives us good reason to think that either 1. or 2. is false, and this is sufficient for showing the safety account's usefulness to process reliabilists.

Now, if our goal in believing is to satisfy (T\*), why might process reliabilism seem attractive? Let's remind ourselves that process reliabilism is the view that

$S$ 's belief that  $p$  is justified *iff* the belief-forming process that caused  $S$  to believe  $p$  at  $t$  is reliable,

where

Belief-forming process  $b$  is reliable *iff*  $b$ 's immediate outputs tend—when employed in a suitable range of circumstances—to yield a balance of true to false belief that is greater than some threshold,  $T$ .

When a belief forming process is reliable, it has a certain modal profile. Namely, its direct outputs are true in a certain range of circumstances. That is, true beliefs that spring from it are not easily false.



Let's suppose that the goal that characterizes the epistemic perspective is that of satisfying (T\*). That a belief is the output of a reliable process will on this account, be a good thing precisely because having beliefs that are not easily false is an end in itself, and the true outputs of reliable belief forming processes are not easily false.

### 3.7.3 Evidentialism and Reliabilism

Reliabilism and evidentialism are often thought of as competitors. Given that the safety account sits nicely with both reliabilism and evidentialism, is this a problem for the view?

I do not think so. As I say earlier in the paper, I do not intend to plump for a particular view of JUSTIFICATION. Rather, I set out to improve and vindicate the thought that belief aims at truth and that (J) follows from this, via means-end rationality.

In addition to giving reliabilists and evidentialists a solid normative foundation, the above may also give us a way to unify these two approaches. From a God's eye point of view, a belief that is true and reliably formed is correct per (T\*). However, from that agent's limited perspective it's possible that reliably-formed true belief seems problematic: to see this, just consider Norman's case. This, I think, shows that there are two perspectives on the goal of satisfying (T\*): an external perspective that looks at a belief's modal profile, and one that looks at it from the point of view of the agent who accepts (T\*). That we can look at beliefs from both perspectives does not make the safety account incoherent, as we're capable of taking both perspectives in many domains.

Now, when (T\*) is the goal, the external perspective seems to coincide with reliabilism and the internal with evidentialism. Given that the former seems to

be a stable externalist position, the latter a stable internalist position, and that the internalist/externalist debate in deadlock, perhaps a view that can accommodate both is called for. The safety view, I propose, can do just that.

### **3.8 Conclusion**

I hope to have shown two things. First, that two tempting instrumentalist approaches to explaining (J) fail. Second, that we can maintain the intuitions that make these views plausible by identifying safety from error as part of the aim of belief.

## Chapter 4

# Transparent Teleology

**Abstract.** According to the teleologist, the fact that there's something wrong with a false belief is explained by the fact that beliefs are guided by a truth-oriented aim, the aim of believing  $p$  only if  $p$ . According to a popular criticism of teleology, the teleologist must choose between two options: guidance by this truth-oriented aim either manifests itself as a disposition to *only* be moved by evidential considerations or a disposition to *sometimes* be moved by evidential considerations. If the former, then the teleologist must deny that truth-insensitive beliefs, like the products or wishful thinking, are beliefs. And, this is clearly implausible. But if teleologists choose the latter option, they cannot explain the fact that in the context of first-personal deliberation about what to believe, we cannot treat non-evidential reasons as reasons to believe. And an account of belief's norms owes us an explanation of this phenomenon. For these reasons—the objection goes—teleology should be rejected. I respond by rejecting both horns of the dilemma.

### 4.1 Introduction

Why is the following a standard by which beliefs are judged?

(T): There is something wrong with false beliefs.

One explanation begins with a claim about the nature of belief. Specifically,

*Teleology:*  $S$  believes  $p$  only if  $p$  was accepted with the aim of thereby accepting a truth.<sup>1</sup>

---

<sup>1</sup>I use the passive voice here intentionally. It might seem more proper to say, " $S$  believes  $p$  only

On this account, beliefs are the products of attempts at accepting a truth. Given this, we can explain why there is something wrong with false beliefs: a false belief is a failure by the lights of the aim with which it was generated. Call this view, the view that takes Teleology to explain (T), *the teleological account*.

The teleological account is, on its face, a promising account of the nature of belief and the norms that apply to it. Beliefs, after all, tend to be sensitive to truth-related considerations. When you ask yourself whether you should believe something—e.g., whether the Democrat will win the next election—you turn to what you take to be the relevant evidence, such as recent polling data. And even if you prefer to believe that the Democrat will win—perhaps on behalf of your mental and physical wellbeing—you can't convince yourself to believe this for these practical reasons. Further, beliefs not formed in conscious deliberation—like perceptual beliefs—seem to display an automatic sensitivity to the truth. If beliefs essentially are attempts at the truth, this is something we would expect. And if this fact about belief can ground (T), the teleological approach gives a parsimonious story about normativity as it applies to belief.

Its initial attractiveness notwithstanding, the teleological account is not a popular view. This is in large part because it is thought to face an insurmountable dilemma. The dilemma runs as follows.

If Teleology explains (T), then either

(s) a mental state is a belief only if it's strongly regulated for the truth,

or

(w) a mental state is a belief only if it's weakly regulated for the truth.

---

if *S* accepted *p* with the aim of thereby accepting a truth." This, however, is misleading since many of our beliefs are spontaneously formed; there are propositions that we believe that *we* do not, in any obvious sense, accept. Perceptual beliefs are an example of this.

In this context a mental state is *strongly regulated for truth* if it is solely regulated by processes that aim at the truth, where a process *aims at the truth* only if it only takes truth-related considerations as inputs.<sup>2</sup> Examples of truth-related considerations include experiences and evidence. For example, if one's perceptual system takes in experiences (and no other inputs) and produces acceptances, it is not ruled out as aiming at the truth.<sup>3</sup> Now consider wishful thinking, which takes one's desires as inputs. Even if wishful thinking takes evidence and experiences as inputs, the fact that it is sensitive to desire rules it out as a process that aims at the truth. Now, suppose Ray's belief as to whether there is a pink-spotted flycatcher in front of him is influenced by wishful thinking and perception.<sup>4</sup> Clearly, Ray's belief is not strongly regulated for truth. But it may be weakly regulated for truth. A mental state is *weakly regulated for truth* if it is regulated by mechanisms that aim at the truth, but also influenced by mechanisms that do not.

The arguments against (s) and (w) run as follows. If (s), then no mental states influenced by cognitive bias are beliefs. But, some mental states—such as Ray's belief about the bird in front of him—count as beliefs and are not strongly regulated for truth. So, (s) is false. But if (w) is true, then the teleological account cannot explain an important observation about belief: we are unable to take practical considerations directly into account when deliberating about what to believe. Further, an account of (T) owes us this. So, (w) is false. So, the teleological account is false.

I proceed as follows. In section 4.2 I give a fuller description of teleology and its competitor, normativism. I then, in section 4.3, present the teleologist's dilemma in greater detail. In section 4.4 I defuse the dilemma.

---

<sup>2</sup>Note that this means that—per Chapter 3—belief must aim not just at truth in the actual world, but non-accidental truth.

<sup>3</sup>I say not ruled out because in presenting the idea of aiming at the truth I only give a necessary condition.

<sup>4</sup>This example is adapted from Feldman (2003).

## 4.2 Belief's Normativity via the Aim of Belief

According to an influential—albeit opaque—adage, “belief aims at the truth” (Williams (1979)). That belief aims at the truth has been taken to explain several commonplaces about belief, a few of which I have already introduced:

**Transparency.** “The deliberative question *whether to believe that p* inevitably gives way to the factual question *whether p*, because the answer to the latter question will determine the answer to the former” (Shah and Velleman (2005)).

(T) There is something wrong with false beliefs.

**Regulation.** Beliefs are weakly regulated for truth.

Williams' adage is clearly a metaphor—beliefs are not little archers with arrows pointed towards the truth (Wedgwood (2002)). To understand how, exactly, belief's aim explains these phenomena, we need to unpack the metaphor. Two prominent accounts of belief's normativity can be understood in terms of how they unpack the metaphor and how they use their account of the adage to explain our commonplaces about belief.

### 4.2.1 Teleology and the Aim of Belief

The teleologist unpacks the metaphor in as literal a way as one sensibly can. She thinks of belief as essentially truth-regulated: part of what it is to be a belief that *p* is to be guided by an aim to accept *p* only if *p*. So “belief aims at the truth” is shorthand for “necessarily, processes that produce beliefs aim to produce true beliefs.” We might now ask, under what conditions does a process aim at the truth?

Before answering this question, we must recognize that there are at least two, and quite different, ways beliefs are formed. One is through first-personal doxastic

deliberation, that is deliberation about what to believe. This is the way we'd ideally manage our belief about the defendant were we on a jury—namely, via conscious consideration of the evidence.

Let me take a moment to briefly describe how I will understand the process of forming a belief via first-personal deliberation. In first-personal doxastic deliberation we ask whether to accept  $p$  and we aim to accept  $p$  only if  $p$  is true. In deliberation we judge—and I understand judging that  $p$  to be a mental act—whether  $p$  is true. Then, if everything goes right, a mental state—the belief that  $p$ —is formed as a result of our having judged that  $p$ .<sup>5</sup>

The other way is fully automatic. When I walk into the courtroom for the first time and see that the defendant is wearing a blazer, there is no deliberation to speak of. There is nothing *I* do to form the belief or judgment that the defendant is wearing a tie. My perceptual faculties spontaneously form the belief that the defendant is wearing a blazer.

In the case of first-personal doxastic deliberation, to say we aim at the truth is to ascribe an intention, desire, or goal at the personal level that guides our deliberation. So, according to the teleological account, one adopts an aim in deliberating about what to believe—an aim that is satisfied only if one ends up accepting a truth at the end of deliberation.

In the case of spontaneously-formed belief, the aim isn't one that's held at the personal level. Rather, the aim is realized by the constitutive aim of the process that forms or regulates the spontaneously formed belief. That is, if a cognitive system was designed—by nature, education, or a designer—to ensure that its acceptances are true.

---

<sup>5</sup>I use the passive voice in this sentence intentionally. I do not think of the process that takes judgments that  $p$  as inputs and has beliefs that  $p$  as outputs as one that we have control over, or—at least—I do not want to think of it as akin to a mental act that *we* perform.

We can now explain how teleologists explain our commonplaces. Transparency is explained by the fact that when one deliberates about whether to believe  $p$ , necessarily, one adopts the intention to believe  $p$  only if  $p$ . So, when one asks herself *whether to believe  $p$* , she looks to considerations that—from her point of view, at least—indicate  $p$ 's truth, i.e., she looks to her evidence. (T) is explained by the fact that forming a belief always involves the possession of an aim, an aim that is frustrated when the belief is false. Lastly, regulation is explained by the fact that an attitude counts as a belief insofar as it is subject to at least some truth regulation. Teleology, then, tidily explains our commonplaces about belief.

#### 4.2.2 Normativism and the Aim of Belief

To deepen our understanding of the teleologist's proposal and to better understand the argument of the next section, it will help to consider the major competitor of the teleological approach, *normativism*. Normativists understand the slogan that belief aims at the truth as *entirely* metaphorical; the metaphor is shorthand for the following:

*Normativism*: it is part of the concept BELIEF that <one ought believe  $p$  only if  $p$ ><sup>6</sup>

The idea behind normativism is that accepting something like (T) is part of what it is to possess the concept BELIEF. On this view, the normative conceptual fact that one ought to believe  $p$  only if  $p$  explains (T).

To bring the contrast between normativism and the teleological approach into full relief, let us see how normativism explains transparency. According to the normativist, part of possessing the concept of belief involves accepting truth as the

---

<sup>6</sup>It's worth noting that the normative content varies from view to view. The debates over the proper formulation of the norm do not affect my arguments. So, I simply use the norm I have stated here as a placeholder for the proper norm, if there is one.



standard for correctness for belief. This functions as a prescription to believe  $p$  only if  $p$ . This, in turn, manifests itself as a disposition to be moved by and only by considerations that one regards as relevant to the truth of  $p$  when deliberating. In other words, in doxastic deliberation, one's reasoning is framed as answering the question *whether to believe  $p$* . So, in doxastic deliberation one's concept of belief is activated. So, in doxastic deliberation one is moved only by considerations she regards as relevant to the truth of  $p$ .

It is worth noting that normativism itself doesn't explain Regulation. On at least one influential version of the proposal—i.e., Shah and Velleman (2005)—we must add to the normativist account the brute stipulation that it's part of the nature of belief to be at least weakly regulated for truth.

In the next section I introduce an influential argument for normativism, The Teleologist's dilemma.

### 4.3 The Teleologist's Dilemma

Shah (2003) uses the phenomenon of transparency to argue against the teleological approach and in favor of normativism. The argument is as follows:<sup>7</sup>

1. Necessarily, first-personal doxastic deliberation exhibits transparency.
2. A mental state can be influenced by non-alethic considerations and count as a belief.
3. (1) cries out for explanation.
4. On the teleological approach, either
  - (s) a mental state is a belief only if it's strongly regulated for the truth, or
  - (w) a mental state is a belief only if it's weakly regulated for the truth.

---

<sup>7</sup>My reconstruction bears a similarity to and was influenced by Mateas (2013)'s.

5. If (s), then (2) is false.
6. So, not-(s).
7. If (w), then (1) is left unexplained.
8. So, the teleological approach can't explain (1).
9. *Assumption*: Normativism (i.e., it is part of the concept BELIEF that <one ought to believe  $p$  iff  $p$ >.)
10. Normativism explains (1) and is compatible with (2)
11. So, probably, Normativism.

Many have taken this argument to deal a decisive blow to teleology while also making a compelling case for normativism. In the next section I explain why this is a mistake.

#### 4.4 Defusing the Dilemma

I do two things in this section. First, I show that even if the teleologist cannot explain transparency, the teleologist's dilemma fails to offer support for normativism. This is because the normativist explanation of transparency, once properly understood, relies on a controversial assumption about motivational internalism. Second, I show that the case against the teleological account should leave us wanting. This is because there are degrees of truth regulation that are strong enough to explain transparency but weak enough to allow cognitively biased attitudes to count as beliefs.

#### 4.4.1 Against (10)

Here, I will explain why—even if we grant (1)-(8)—the teleologist’s dilemma does not make a compelling case in favor of normativism.<sup>8</sup>

The main reason to be suspicious of this argument is that (9) assumes an implausibly strong form of motivational internalism. Recall that crucial to the normativist’s account is something like this:

13. Part of possessing the concept of belief involves accepting truth as the standard for correctness for belief.
  - (a) This functions as a prescription to believe  $p$  only if  $p$ .
  - (b) This manifests itself as a strong disposition to be moved by only by considerations that one regards as relevant to the truth of  $p$  when deliberating.

Clearly, the claim that accepting a prescription to  $\phi$  manifests as itself as a strong disposition to  $\phi$  when whenever the question of whether to  $\phi$  arises plays a large role in the normativist’s account of transparency.

But this claim—(13b)—is highly suspect. It’s an unfortunate fact of our agency that we often accept—and fully endorse—a prescription and yet fail to follow it, either because of akrasia or incompetence. Once we see this, we should see that we have little reason to accept (13b). Thus, we have little reason to think normativism explains transparency.

The normativist may respond by saying that this is too quick. The claim under discussion here doesn’t rely on the assumption that we’re always strongly motivated to follow prescriptions we endorse, she might say. Rather, when accepting a prescription is constitutive of a concept—as it is in the case of belief—we display this strong motivation.

---

<sup>8</sup>For a similar arguments see Steglich-Petersen (2006), Mateas (2013), and Mchugh (2011).

I don't think that we should be convinced by this caveat. To see why, consider an example. Suppose that knowledge is the norm of assertion. Or, more precisely, assume:

(KNOW) One ought: assert that  $p$  only if one knows that  $p$ .

Let us suppose that part of possessing the concept ASSERTION involves accepting that knowledge as the norm of assertion. Were this the case, it seems that we could intentionally violate it. That is, it seems that under these assumptions we could, for example, tell lies. And for this reason, it seems that even a more cautious version of the assumption that under-girds (13) is false, or—at best—in need of further defense. For this reason, I do not think that Shah's dilemma is a convincing argument for normativism, even if we grant (1)-(8).

Now, this discussion may seem question-begging. The claim under consideration is whether norms that are constitutive of concepts instill in us strong dispositions to satisfy those norms. In the above case, I give an example of an alleged constitutive norm and then showed that it may not be motivating. Perhaps, the normativist might say, this only serves to show that (KNOW) is not a constitutive norm of ASSERTION.

I don't think that this objection quite works. The prescription <believe  $p$  only if  $p$ > may not be a constitutive norm either; this, of course, is one of the central claims of this paper. What is important is when we look to plausible instances of constitutive norms, we do not see the strong motivation internalism associated with the normativist account of belief. What is helpful about this example is that while it is uncontroversial that we can speak at will; it is also uncontroversial that whatever  $\phi$ 'ing at will amounts to in this context, we *cannot* believe at will. Examining these cases side by side helps us to see that the normativist explanation

may only seem plausible in the case of belief because belief is, unlike speaking, not something that we have voluntaristic control over.

This, I think, shows that the normativist lacks the clear dialectical advantage over teleologists that they are sometimes thought to have. It is also worth noting that this is not the only problem that normativism suffers from in the context of their abductive argument against the teleological account.

Abductive arguments, of which the teleologist's dilemma is an instance, work by showing that for some set of observations, one hypothesis better explains those observations than some other hypothesis. I have just shown that we should lack confidence that the normativist explains what she purports to explain. This is a weakness in the argument. Another weakness is that a crucial observation about belief is left out of the observation set: Regulation. Normativism does not even purport to explain this phenomenon. Teleology, on the other hand, does.

I think that these considerations show, at the very least, that the teleologist's dilemma fails to give a convincing argument in favor of normativism. However, showing that the dilemma fails to give a convincing argument against normativism doesn't suffice to show that the teleologist is scot-free. (1)-(8) constitute a powerful challenge to teleology. In what remains of this section I will show that this objection can be overcome, because (5) and (7) are false. While showing either of these premises to be false is sufficient for overcoming the dilemma as stated, showing both to be false is not redundant. This is because the rejection of both premises play an important role in developing the path forward that I will suggest for teleologists.

#### 4.4.2 Against (5)

I will begin with the case against (5), the premise which claimed that if it's part of the nature of belief to be strongly regulated for truth-related considerations, then it's false that a mental state can be influenced by non-alethic considerations and count as a belief. I will then deliver the case against (7), the premise which claims that if it isn't part of the nature of belief to be weakly regulated for truth-related considerations, then transparency is left unexplained.

I will begin my case by making a few distinctions. The teleologist's dilemma makes use of the distinction between strong and weak truth regulation. Recall that to be weakly truth-regulated is to be sensitive to alethic considerations but influenced by other factors. To be strongly truth-regulated is to be sensitive only to alethic considerations. There are, as I will soon show, several ways in which we can conceive of both strong and weak truth regulation. Once we see this, as I will argue, we can identify degrees of truth-regulation that can both explain transparency *and* allow attitudes influenced by non-alethic considerations to qualify as beliefs.

##### 4.4.2.1 Varieties of Strong Truth-Regulation

Let us begin our discussion, then, by identifying some varieties of strong truth-regulation. There is a debate as to whether

**Uniqueness:** a total body of evidence justifies at most one doxastic attitude towards any particular proposition.

I will not engage with arguments for or against Uniqueness here—my arguments do not depend on its truth in any way. But let us assume, for a moment, that Uniqueness is true. What would this amount to? One thing the truth of uniqueness would amount to is there being *one* set of belief regulation policies that it is correct

to abide by and for those policies to always give a determinate (though not always determinable) answer to the question *what should I believe*. It would also imply that there is *one* correct mapping of batches of evidence (the inputs) onto acceptance-states (the outputs). Call the set of policies that it is correct to abide by if Uniqueness is true *U*.

Now, consider all of the possible sets of evidential policies one could have. Some of these policies are in tension with *very* low standards that we might want our belief regulation policies to meet, given the adoption of the truth aim. For example, some possible—but problematic—standards will recommend that in some circumstance we believe *p* and believe not-*p*. Clearly, this is problematic. Why? Well, one thing teleologists can say is that evidential policies that lead agents to believe in contradictions are evidential policies that, with certainty, lead us astray from truth. Now, note that the norm “don’t believe contradictions”—while a meaningful constraint on which attitudes we ought to have—is a merely *formal* constraint on what to believe. That is, it is a norm of how our attitudes—whatever their contents may be—ought to be patterned. At *minimum*, we want norms of belief regulation to embody certain formal constraints, like coherence. Call any set of belief regulation policies that satisfy the right formal constraints—whatever those constraints turn out to be—*formally rational*.

Let us also say that *U* defines the strongest form of truth regulation; that is, let us say that the strongest form of truth-regulation that the teleologist might avail herself to is

*U-Regulation*: *S* believes *p* only if her attitude towards *p* is regulated by a process that flawlessly enforces *U*.

While *U-Regulation* is the strongest form of truth regulation, we shouldn’t assume that managing one’s beliefs in accordance with *U* is the *only* way for one to strongly

regulate her doxastic attitudes for truth. Any set of policies that is formally rational requires one to manage her beliefs in ways that are sensitive to only to the evidence; all of these policies are mappings of batches of evidence onto acceptance-states. And so managing one's beliefs in accordance with a formally rational set of standards other than  $U$  can constitute strong truth-regulation.

Let's pause, for a moment, to explain why reasoning in accordance with any formally rational standards—a set which includes  $U$ —will exhibit transparency. As I say above, all of these policies are mappings of batches of evidence onto acceptance-states. Let's imagine that reasoning works the following way. Suppose every possible batch of (total) evidence you could have has a name  $e_1 \dots e_n$ . Further, suppose that  $U$  (as well as any of the formally rational standards) just is a long list of instructions:

If  $e_1 \rightarrow$  you may accept propositions in  $\mathbb{B}_1$ , may not accept propositions in  $\mathbb{D}_1 \dots$ <sup>9</sup>

If  $e_2 \rightarrow$  you may accept propositions in  $\mathbb{B}_2$ , may not accept propositions in  $\mathbb{D}_2 \dots$

...

If  $e_n \rightarrow$  you may accept propositions in  $\mathbb{B}_n$ , may not accept propositions in  $\mathbb{D}_n \dots$

Now, imagine you have batch of evidence that you correctly identify as  $e_4$ . What do you do? You pick up your instructions—those which encode  $U$ , say—and thereby accept (and reject, suspend judgment over, take .25 credence towards, etc.) propositions accordingly. While reasoning, no doubt, does not work this way, we can model it as if it did.

Among the formally rational standards, some might be correct and others not. Those who deny Uniqueness are split on whether all formally rational standards are permissible or whether only some are. Call those who think *some* formally

---

<sup>9</sup> $\mathbb{B}_1 \dots \mathbb{B}_n$  and  $\mathbb{D}_1 \dots \mathbb{D}_n$  are sets of propositions.



rational standards are permissible *modest permissivists*, and call those who think that *all* are permissible *extreme permissivists*. Some formally rational standards are strange enough that some permissivists shy away from identifying them as permissible. These permissivists are the modest ones, they think there's more than one permissible set of belief regulation policies but they don't think anything goes with respect to the formally rational standards.

It is worth noting that we can now define two more kinds of truth regulation that each count as strong truth regulation. Call the set of standards extreme permissivists approve of *E*. Similarly, call the set of standards modest permissivists approve of *M*. If modest permissivism is true, then any agent who flawlessly regulates her beliefs in accordance with a standard in *M* strongly regulates her beliefs for truth. Similarly, if we accept extreme permissivism is true, then any agent who flawlessly regulates her beliefs in accordance with a standard in *E* strongly regulates her beliefs for truth.

#### 4.4.2.2 Strong Truth-Regulation, Cognitive Bias, and the Shotgun Approach

In what follows I demonstrate several ways in which cognitive bias is compatible with each of the degrees of strong truth-regulation described in the previous section. I do not endorse *any* of these as the correct degree of truth-regulation for teleologists to adopt. Rather, I think that teleologists can and should accept something weaker. I will describe this degree of truth regulation in 4.4.3.

Why, then, is this discussion necessary? The primary reason is that the teleologist's dilemma implies that *any* level of truth-regulation strong enough to explain transparency will admit of too few beliefs. Let's call this the *too few beliefs problem*. In this section I identify a variety of types of thinking that are biased by non-alethic considerations that are compatible with various strong forms of strong truth regu-

lation. In doing this, I lay the groundwork for my account's ability to deal with the too few beliefs problem. This is important because my strategy for dealing with the too few beliefs charge will be to take a shotgun approach: I won't give any *one* explanation for how cognitive bias is compatible with a degree of truth regulation strong enough to explain transparency; rather, I will show that there are a variety of strategies for explaining the compatibility in each local case.

#### 4.4.2.2.1 *U* & Biased Thinking

Let us suppose

*U-Regulation*: *S* believes *p* only if *S*'s belief that *p* is solely influenced by a process that flawlessly enforces *U*.

On this picture, to fail to regulate one's (would-be) beliefs in accordance with *U* amounts to a failure to be sensitive only to evidence, and a failure to believe at all. This is true even if one has adopted a set of standards in *M* that isn't *U* and perfectly abides by those standards.

An account that takes this to be the correct account of truth regulation necessary for the formation of belief is clearly able to explain transparency. If forming a belief via first-personal deliberation requires successfully abiding by *U*, then whenever one engages in first-personal doxastic deliberation, one is only sensitive to evidential considerations. And if (5) is true, this account is incompatible with *any* mental attitude influenced by cognitive bias counting as a belief.

But, it's not the case that a belief can't be both strongly regulated for truth and influenced by cognitive bias. So see this, just consider the following case:

**Confirmation Bias.** I want *A* to be true, so I seek evidence that confirms *A*. I am alienated enough from what is going on that I end up with a batch of evidence,  $e_a$ , that, per *U*, supports *A*. Using *U* I come to believe *A* on the basis of  $e_a$ .

Certainly, cases like Confirmation Bias are possible. In this story, my belief that  $A$  is strongly truth regulated—the process that generated it, reasoning, only took  $e_a$  as its input. And yet my belief was subject to the influences of a cognitive bias, confirmation bias. If Confirmation Bias isn't compelling, let's consider one more example.

**Egocentric Forgetting.** For my own wellbeing, it would be best if I remember clearly some successes and forget the pains of certain failures. Lucky for me, my brain is wired to engage in egocentric forgetting; that is, I am wired to forget (remember) such details. (And I'm wired this way so as to promote my wellbeing.) I fail miserably at public speaking on an occasion,  $o$ . Soon after, egocentric forgetting permanently erases the worst memories I have of  $o$ . Were I to remember all of the details of  $o$ , I would be required—relative to  $U$ —to believe <I am a bad public speaker>. But my evidence has been manicured through egocentric forgetting, and the evidence I have, per  $U$ , supports <I am a decent public speaker>. I update on my evidence accordingly.

Egocentric Forgetting is a bit fantastical, but it's a perfectly coherent story. In any event, it serves to give second example where strong regulation and cognitive bias aren't mutually exclusive.

Let me briefly describe what our two cases have in common as a means of identifying, in general, one way in which our beliefs can be biased and strongly truth regulated. In both cases there is an interest—in one case the interest is based in desire, in the other wellbeing—in believing a proposition. The interest works as a filter for what goes into reasoning. But, importantly, even though filtering has occurred, everything that goes into reasoning is evidence. Further, the interest doesn't cause reasoning to misfire—reasoning does not take the interest itself as an input, nor does it improperly weight the evidence that it is fed. In any case like this, the result is belief that is influenced by bias, but belief that is strongly regulated for truth.

If I've just shown (5) to be false—and I think I have—we cannot rest content. We have yet to overcome the real challenge posed by the teleologist's dilemma. Presumably, one can fail to abide by  $U$  in deliberation and yet still form a belief as a result of that deliberation. The present account denies this, and so admits of too few beliefs. For this reason I will, in the next section, discuss a slightly weaker standard.

#### 4.4.2.2 Formally Rational Standards & Wishful Thinking

Let us now drop the assumption that only  $U$  is correct. Suppose now that

*P-Regulation: S believes  $p$  only if  $S$ 's belief that  $p$  is solely influenced by processes that flawlessly enforces a standard in  $M$ .*

Under the assumption that modest permissivists are correct, two new forms of wishful thinking compatible with transparency emerge.

Both forms have to do with the fact that permissivism opens the possibility of two agents correctly assessing the evidence, relative to the standards each of them accepts, and disagreeing about what the evidence says. For instance:

Suppose that six months before the US presidential election, it is quite unclear whether the Democratic or the Republican nominee will win. (Although it is clear that one or the other will.) I possess a large body of information that I take to bear on this question. Some of this information makes it more likely that the Democrat will win, while some of it makes that outcome less likely. On balance, I regard it as somewhat more likely that the Democrat will win than not, so I invest somewhat more credence in that proposition than in its negation. [...] Suppose [...] that you and I agree on the basis of our common evidence that the Democrat is more likely than not to be elected. We similarly agree that although this outcome is more likely than the alternative, it's far from a sure thing. The only difference between us is this: you're a bit more cautious about the Democrat's prospects, and so give a bit less credence to the proposition that the Democrat will win than I do. Here there seems

little pressure for me to conclude that you are less reasonable than I am. Moreover, the natural verdict about the case is that it's consistent with everything that's been stipulated so far that you and I might both be fully reasonable in our opinions about the election, despite the fact that those opinions are not identical (Kelly 2013, pp. 299–300).

By stipulation, there's no evidence for the underlying policy that leads you to be cautious of the Democrat's chances, nor is there evidence for the underlying policy that leads me to be cavalier about her chances. Now, suppose that I'm a Democrat and you're a Republican. Further, suppose we are both hedonists who have deliberately adopted the standards we accept for this reason. Under these assumptions, our beliefs regarding the election outcomes are both wishful and strongly truth-regulated.

The second form builds on the first. Suppose I drift from one set of permissible standards to another, and for practical reasons. Suppose that I really want to do away with the stress of thinking the Republican might win. Suppose further that I fully understand some set of standards in  $M$  that's even more optimistic with respect to the Democrat's chances than my current standards are. Finally, suppose that because of this I adopt this new set of standards and bring my attitudes in line with them. My new attitude towards the candidate is presumably an instance of wishful thinking. But it correctly reflects the standards that I presently accept, and, so, is strongly truth-regulated.

While both cases, again, are fantastical, they demonstrate a new way in which our interests might bias our beliefs. In both cases we can think of our interests as intervening on our standards. The interests, unlike the cases of the previous section, don't *filter* our evidence. Rather they influence which function in  $M$  we deploy. Note, however, that when we deploy an  $M$ -function, our deliberation will be characterized by transparency, will only take evidence as inputs, and will reflect

a correct interpretation of the evidence.

To close this section, let's *briefly* consider standards in *E* but not *M*. Some of these standards will treat certain forms of wishful thinking as legitimate ways to manage one's attitudes. And so, *if* extreme permissivism is correct, then, relative to certain policies, one's desires *are* alethic considerations; so, there will be an increase in cases where one can rationally and wishfully believe. And yet, for the above stated reasons, one's deliberation will be characterized by transparency, will only take evidence as inputs, and will reflect a correct interpretation of the evidence.

#### 4.4.2.3 Earnest Misapplication of Correct Standards

Let us suppose that I endorse *U*, that when I deliberate about what to believe I *try* to believe in accordance with *U*, and when I do this I only consider the evidence. Suppose that I do this by adjusting my beliefs to what I *take* the evidence to say, but that I don't always get things right.

We can define a degree of truth regulation that is quite weak and yet compatible with transparency:

*S-Regulation*<sub>1</sub>: *S* believes *p* only if *S*'s belief that *p* is solely influenced by processes that earnestly enforce correct evidential standards.

To make vivid the kind of mistake that this weak standard is compatible with, consider the following sort of case. Imagine that I am considering whether to give a public lecture. I have no egocentric forgetting module as described before. Rather, in deliberating as to whether to give a talk, I erroneously turn my mind's eye to a subset of the evidence that I have—the evidence that supports <I am a great public speaker>. In that moment I earnestly, but mistakenly, ignore much of the evidence that I have. Suppose further that the batch of evidence I focus on

would support <I am a great public speaker> were it in fact my total evidence. However, my total evidence supports the negation of this proposition.

This account, again, predicts transparency. This is because on this account, if I form a belief through first-personal deliberation, I do so by apportioning my belief the evidence and nothing else. In other words, from my point of view, the question *whether to believe p* is answered by the answer to the question *whether p is true*; I'm just responding to the evidence incorrectly.

#### 4.4.2.4 Earnest Misapplication of Incorrect Standards

There is an even weaker strong standard that predicts transparency. Suppose, for example, that uniqueness is true and I have internalized one of the (non-*U*) *M* standards. If I earnestly—and incorrectly—apply those standards to my evidence, the question *whether to believe p* will be answered by what I *take* to be the answer to the question *whether p is true*. This means we can accept a standard even weaker than *S-Regulation*<sub>1</sub> and account for transparency. That is, we can accept

*S-Regulation*<sub>2</sub>: *S* believes *p* only if *S*'s belief that *p* is solely influenced by processes that earnestly enforce evidential standards.

In virtue of the many types of cognitive bias that it is compatible with, *S-Regulation*<sub>2</sub> defines a very weak degree of truth regulation. However, the teleologist cannot appeal to it in order to successfully overcome the too few beliefs problem. This is because *W-Regulation*<sub>2</sub> is, in a sense, still too strong. For consider:

**Eyes-Off.** I have a good understanding of what *U* requires, and *in deliberation* I am very good at bringing my attitudes in line with *U*. At *t*<sub>1</sub> I see that my evidence supports <the Democrat will win>, and, so, I believe accordingly. But a moment later, when my mind's eye is not focused on that belief, I give in to pessimism and become skeptical that

the Democrat will win. We can imagine that whenever I become aware of the fact that I'm sliding into pessimism, I consider the evidence and—via first-personal deliberation—bring my confidence in the proposition back to where it should be, according to *U*'s dictates.

Qua *S-Regulation*<sub>2</sub> whatever mental attitude I have towards <the Democrat will win>, it can't be belief. An this is clearly unacceptable—Eyes-Off is quite obviously compatible with my mental attitude counting as belief. So, we need something even weaker.

#### 4.4.2.5 General Remarks

So far, we have considered standards that count as strong truth regulation. To accommodate Eyes-Off, we not only need a standard *weaker* than *S-Regulation*<sub>2</sub>, we need one that is *weak*, full stop. This means we need to change course; we need a discussion of weak standards, and we need to explain how those standards predict transparency. I will do this in the next section.

Before doing this, I want to briefly take stock of what we have learned in this section. First, we have learned that transparency is predicted by fairly weak standards. This is important because in order to overcome to the too few beliefs problem, we need an account of belief that is compatible with any number of cognitive biases. The discussion above is meant to shed light on the different ways in which a standard of regulation strong enough to predict transparency might also be compatible with a wide class of biased beliefs.

In the next section, I will discuss a standard that is weak, full stop. This standard is weaker than *S-Regulation*<sub>2</sub>. So if it predicts transparency, it inherits all of the explanatory power of *S-Regulation*<sub>2</sub>.



### 4.4.3 Against (7)

I now turn to a discussion of weak truth regulation. On any proposal considered in the last section, agents believe only if they only believe solely on the evidence. How might we relax that assumption while accounting for transparency? In this section I discuss one way we can accomplish this. The key is to focus on a crucial claim of weak truth regulation: that being the product of a process that aims at the truth is sufficient for being a belief.

#### 4.4.3.1 Weak Truth Regulation and Transparency

The previous sections assumed that a mental state is a belief only if it is strongly regulated for truth. This assumption, as I have demonstrated, is flawed; but, it made accounting for transparency easy. How might we recover this?

Let us begin by defining a weak standard:

*W-Regulation:* *S* believes *p* only if *S*'s mental state that *p* was generated by or is regulated by a process that—when deployed—earnestly enforces an evidential standard.

Note that on this account, my mental state in Eyes-Off is not ruled out from being a belief.

Now we ask: how is an account that accepts W-Regulation going to predict transparency?

On the W-Regulation picture, each belief is a belief at least partially in virtue of the fact that some process that regulates it aims at the truth, in the sense that the process earnestly enforces an evidential standard. Doxastic deliberation is one such process—per Shah, one engages in doxastic deliberation only if one only takes into consideration what one takes to be the evidence. And so—given the plausible

assumption that doxastic deliberation aims at the truth—the teleologist can explain transparency. The arguments from the last section show that this is compatible with a wide number of cognitive biases. Further, if we accept W-Regulation, it's compatible with the mental state in Eyes-Off counting as a belief.

There might be a lingering concern. One might think: W-Regulation is compatible with my desiring  $p$ , accepting  $p$  because  $p$  is desirable, and then believing that  $p$  in virtue of inferring  $\langle p \rangle$  from  $\langle \text{it's desirable that } p \rangle$ . But, this is not quite right. It *is* right that *W-Regulation* is compatible with one's desiring  $p$ , accepting  $p$  because  $p$  is desirable, and then believing that  $p$ . But it is *not* the case that one can believe  $p$  in virtue of of this. This is because the teleologist accepts something like the following

**Claim.** If a line of reasoning issues a mental state and that mental state is a belief *in virtue of the nature of that line of reasoning*, then the line of reasoning was an inquiry as to whether  $p$ .

Once we see this, we can see why the charge that teleology is unable to explain transparency was taken to have such force. The dilemma may make it sound as though the proponent of weak regulation must deny Claim. But this is false. Claim is a very important part of the picture; it explains why it's the case that, necessarily, if one's line of deliberation makes it the case that one believes  $p$ , in that line of deliberation one has considered *whether*  $p$ .

## 4.5 Conclusion

According to the teleologist's dilemma, the teleologist must choose between two options: strong or weak truth-regulation. If the former, then the teleologist must

deny that truth-insensitive beliefs—like the products or wishful thinking—are beliefs. And the products of wishful thinking clearly are beliefs. But if the teleologist chooses latter option, her account cannot explain the fact that in the context of first-personal deliberation about what to believe, we cannot treat non-evidential reasons as reasons to believe. And an account of belief's norms owes us an explanation of this phenomenon. And so, teleology must be rejected. Further, normativism *can* explain transparency without denying that wishful thoughts can be beliefs. So, we should be normativists.

I take myself to have shown that these conclusions do not follow. Three crucial claims of the dilemma are problematic. That strong truth-regulation and wishful thinking are incompatible was shown to be false in 4.4.2. That weak truth-regulation cannot explain transparency was shown to be false in 4.4.3. And even if the arguments of 4.4.2 and 4.4.3 are mistaken, we still have good reason to resist to push towards normativism. This is because—as I demonstrated in 4.4.1—normativism's explanation of transparency does not explain the fact that beliefs are weakly truth regulated; it also relies on a highly controversial claim about motivational internalism.

# Bibliography

- Ahlstrom-Vij, K. and Dunn, J. (2014). A defence of epistemic consequentialism. *Philosophical Quarterly*, 64(257):541–551.
- Alston, W. (1989). *Epistemic Justification*. Cornell University Press.
- Alston, W. (2005). *Beyond Justification: Dimensions of Epistemic Evaluation*. Ithaca: Cornell University Press.
- Alston, W. P. (1985). Concepts of epistemic justification. *The Monist*, 68(2):57–89.
- Berker, S. (2013a). Epistemic teleology and the separateness of propositions. *Philosophical Review*, 122(3):337–393.
- Berker, S. (2013b). The rejection of epistemic consequentialism. *Philosophical Issues*, 23(1):363–387.
- Bonjour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, 5(1):53–73.
- Bonjour, L. (1985). *The Structure of Empirical Knowledge*. Harvard University Press.
- Brandt, R. B. (1992). *Morality, Utilitarianism, and Rights*. Cambridge University Press.

- Côté-Bouchard, C. (2015). Epistemic instrumentalism and the too few reasons objection. *International Journal of Philosophical Studies*, 23(3):337–355.
- Côté-Bouchard, C. (forthcoming). Ought implies can against epistemic deontology: Beyond doxastic involuntarism. *Synthese*, pages 1–16.
- Fassio, D. (2015). Belief, aim of. In *Internet Encyclopedia of Philosophy*.
- Feldman, R. (2003). *Epistemology*. Foundations of Philosophy Series. Prentice Hall.
- Foley, R. (1987). *The Theory of Epistemic Rationality*. Harvard University Press.
- Gensler, H. J. (2013). *Ethics and the Golden Rule*. Routledge.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.
- Gibbons, J. (2013). *The Norm of Belief*. Oxford University Press.
- Haji, I. (2002). *Deontic Morality and Control*. Cambridge University Press.
- Hazlett, A. (2013). *A Luxury of the Understanding: On the Value of True Belief*. Oxford University Press.
- Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, 66(3):612–640.
- Kornblith, H. (1993). Epistemic normativity. *Synthese*, 94(3):357–376.
- Leite, A. (2007). Epistemic instrumentalism and reasons for belief: A reply to tom kelly’s “epistemic rationality as instrumental rationality: A critique”. *Philosophy and Phenomenological Research*, 75(2):456–464.

- Levy, N. (2007). The responsibility of the psychopath revisited. *Philosophy, Psychiatry, and Psychology*, 14(2):129–138.
- Littlejohn, C. (2012). Does ‘ought’ still imply ‘can’? *Philosophia*, 40(4):821–828.
- Maguire, B. and Woods, J. (Manuscript). Explaining epistemic normativity.
- Mateas, J. T. (2013). Is there an ‘ought’ in belief? *Teorema: International Journal of Philosophy*, 32(3):75–90.
- Mchugh, C. (2011). What do we aim at when we believe? *Dialectica*, 65(3):369–392.
- Owens, D. J. (2003). Does belief have an aim? *Philosophical Studies*, 115(3):283–305.
- Papineau, D. (2013). There are no norms of belief. In Chan, T., editor, *The Aim of Belief*.
- Schroeder, M. A. (2007). *Slaves of the Passions*. Oxford University Press.
- Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford University Press.
- Shah, N. (2003). How truth governs belief. *Philosophical Review*, 112(4):447–482.
- Shah, N. and Velleman, J. D. (2005). Doxastic deliberation. *Philosophical Review*, 114(4):497–534.
- Sharadin, N. (forthcoming). Epistemic instrumentalism and the reason to believe in accord with the evidence. *Synthese*, pages 1–19.
- Sidgwick, H. (1901). *Methods of Ethics*. Kaplan.
- Smart, J. J. C. (1956). Extreme and restricted utilitarianism. *Philosophical Quarterly*, 6(25):344–354.

- Steglich-Petersen, A. (2006). No norm needed: On the aim of belief. *Philosophical Quarterly*, 56(225):499–516.
- Street, S. (2009). In defense of future tuesday indifference: Ideally coherent eccentrics and the contingency of what matters. *Philosophical Issues*, 19(1):273–298.
- Streumer, B. (2003). Does ‘ought’ conversationally implicate ‘can’? *European Journal of Philosophy*, 11(2):219–228.
- Velleman, D. (2000). On the aim of belief. In Velleman, D., editor, *The Possibility of Practical Reason*, pages 244–81. Oxford University Press.
- Vranas, P. B. M. (2007). I ought, therefore I can. *Philosophical Studies*, 136(2):167–216.
- Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 36(s16):267–97.
- Williams, B. (1979). Internal and external reasons. In Harrison, R., editor, *Rational Action*, pages 101–113. Cambridge University Press.
- Zimmerman, M. J. (1996). *The Concept of Moral Obligation*. Cambridge University Press.