Title
The Ethics of Self-Fulfilling Belief

Permalink
https://escholarship.org/uc/item/2f2597vr

Author
Antill, Gregory Elias

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Ethics of Self-Fulfilling Belief

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Philosophy

by

Gregory Elias Antill

2016

ABSTRACT OF THE THESIS

The Ethics of Self-Fulfilling Belief

by

Gregory Elias Antill

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2016

Professor Pamela Hieronymi, Chair

This dissertation examines how we ought to reason about propositions whose truth is determined by whether we believe them. In it, I defend the thesis that in cases of self-fulfilling belief we ought to believe whatever would be best, if true. Though believing whatever would be best if true appears to be a form of wishful thinking, and so unwarranted, this dissertation develops an account according to which, when a belief is self-fulfilling, *optimistic reasons* which show what we believe to be good can also be genuine epistemic reasons for believing it to be true.

The dissertation of Gregory Elias Antill is approved.

Calvin G. Normore

Eric Schwitzgabel

Pamela Hieronymi, Committee Chair
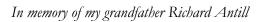
University of California, Los Angeles
2016

In memory of my grandfather Richard Antill

The Ethics of Self-Fulfilling Belief

Table of Contents

Education

University of California, Los Angeles
M.A In Philosophy, 2011

Brandeis University
Bachelor of Arts in History and Philosophy, summa cum laude, 2009

Introduction

This dissertation concerns the nature and rationality of self-fulfilling beliefs: beliefs whose contents will be true just in case you believe them, because you believe them. Examples of this phenomenon span the quotidian – a child's belief that she will be fed may prompt a parent to begin her feeding – to the complex – as in cases, from the psychology of education, in which student performances match the expectations of their instructors.

These examples can be difficult to fit into traditional theories of theoretical reasoning, where the role of theoretical reasoning is to get us on to some independent fact of the matter, by following our evidence. Since there is no independent fact of the matter to track when a belief is self-fulfilling, there will be no evidence of that fact for us to follow. But we are not Cartesian egos, apart from the world and observing it. We ourselves are part of the world we are trying to represent, and so, sometimes, what we believe can affect what the objective world is like. We need an account of theoretical reasoning which can accommodate this fact, and explain how we ought to deliberate about those states of affairs effected by our deliberating.

This dissertation defends the thesis that when a belief is self-fulfilling, we ought to believe whatever would be best, if true. The child ought to believe she will be fed because she is hungry; the teacher ought to believe that her students will perform well on the test because a good performance will make their lives go better. Many philosophers have been skeptical of such an account because it seems to require accepting the existence of pragmatic reasons for belief. That is, it seems to require accepting that our beliefs can be justified in virtue of considerations that show the belief useful, rather than true. I am also skeptical of such pragmatic reasons. Rather than argue for the existence of pragmatic reasons for belief, I argue that failing to believe what would be best, if true, is an *epistemic* failing.

In this dissertation, I will develop an account of the nature of self-fulfilling belief according to which when a belief is self-fulfilling, *optimistic reasons,* which show what we believe to be good, can also be genuine epistemic reasons, which show what we believe to be true. I argue that self-fulfilling beliefs afford us a special way of engaging in theoretical deliberation about their content: rather than discovering what is true on the basis of evidence, we can instead decide what is to be true on the basis of optimistic reasons.

In *chapter one*, I discuss the nature of self-fulfilling belief. I consider objections to the possibility of self-fulfilling beliefs, on the grounds that self-fulfilling beliefs have the wrong direction of fit: beliefs, to be *beliefs,* must depend on the facts and not the other way around. I argue that the force of such objections relies upon a failure to distinguish between the relevant senses of 'dependence' at issue. What matters for belief is that its correctness depend on whether the believed state of affairs obtains; but this is consistent with that state of affairs being caused, or generated, by the belief.

In *chapter two*, I consider the epistemic status of self-fulfilling beliefs. Self-fulfilling beliefs may appear to be in especially good epistemic shape. If an agent knows her belief that p is self-fulfilling, she will have good evidence that her belief is true. I argue that an important and under-appreciated epistemic puzzle arises if we focus on the perspective of the agent reasoning about whether to adopt the belief in the first place: because the agent will be correct no matter what she believes, she has no reason to believe, rather than disbelieve, that the proposition in question is true.

In *chapter three*, I criticize the view in the literature that self-fulfilling beliefs constitute a special case in which we can permissibly believe for pragmatic reasons, because whatever we choose to believe will end up true. I argue that this view fails to distinguish between the aim of acquiring a true belief and the aim of believing what is true. While one cannot usually

fail to establish that one will acquire a true belief without establishing the truth of the believed proposition, in the case self-fulfilling belief the two can come apart. I argue that insofar as the aim of belief has to do with determining whether the believed proposition is true, it will be both impossible and impermissible to believe for pragmatic reasons.

In *chapter four*, I advance an original account of how believing for optimistic reasons can provide epistemic warrant for a belief. In the first part of chapter four, I distinguish pragmatic reasons, which show something good about having a belief, from optimistic reasons, which show something good about what we believe. In the second part of chapter four I argue that, unlike pragmatic reasons, optimistic reasons can be the proper grounds for a distinctive form of reliable, truth-directed, inference: determining what is true by deciding what is to be true. Since optimistic reasons, so employed, help a believer determine whether her belief is true, these optimistic reasons can be genuine epistemic reasons for that belief.

In *Chapter five,* I conclude the dissertation by applying the theory of optimistic reasons, developed in chapter four, to the case of self-fulfilling belief: when a belief is self-fulfilling, we come to believe that p by deciding that p be so. Arguing this involves establishing that self-fulfilling beliefs are a special instance of the larger phenomenon of our capacity to believe for optimistic reasons in cases where we have control over the states of affairs we are reasoning about. When we see that what happens depends on what we believe will happen, we are in the position of deciding what is to happen, on the basis of the considerations which show it ought to.

Chapter 1: Self-Fulfilling Beliefs

In this chapter, I hope to provide some groundwork for further discussion by exploring the nature of self-fulfilling beliefs. I will start by describing in greater detail what it is for a belief to be self-fulfilling, and a challenge to the notion that self-fulfilling beliefs, so described, can really be *beliefs,* on the grounds that such beliefs would have the wrong direction of fit. I will then survey the various ways in which beliefs might *be* self-fulfilling, drawing out some commonalities of self-fulfilling belief which will be important in future discussions of those beliefs' epistemological import.

§I. What is a Self-Fulfilling Belief

Formally, as I will understand the phenomenon of self-fulfilling belief, for a belief to be self-fulfilling is for that belief to bear the following polyadic relationship: a belief, B, in some proposition, p, is self-fulfilling for some subject, S, in a context, C, just in case if S believes that p in C, then p, where p is true in virtue of S's belief that p in C.

I will begin by unpacking this formulation. Fist, as I will understand it, the property of being self-fulfilling is property enjoyed by propositional beliefs, like the belief that I will succeed or the belief that Jones will like me. The property of being self-fulfilling is not a property of "objectual beliefs" like a belief in God, or in Jones, or in yourself, where this is a matter of your having some faith, or trust, in the object you believe in, rather than believing that the object exists, or that the object is trustworthy.[1] A belief is self-fulfilling when the proposition which the belief is an attitude toward is true, and true in virtue of the belief in that proposition.

Second, self-fulfilling beliefs have a *relational character.* This is because the propositions made true will, in most circumstances, be made true by particular token

---

[1]See Audi (2008) or Alston (1996)

instances of belief, and so a belief that p will be self-fulfilling only relative to certain believers

in a certain circumstances. It may be that I, in believing that I will win the match, will make it

the case that I win the match, but that you, in believing that I will win the match, will not

make it the case that I win the match. Whether a belief is self-fulfilling is thus partly a matter

of whose belief it is. Moreover, it may be that if I believe that I will win the match while

playing, I will make it the case that I will win the match, but that my belief that I will win the

match hours before (or hours later) will not have any effect on whether I win the match.

Whether a belief is self-fulfilling is thus a matter, not just of who is believing, but when, and

in what context, they have the belief.[2]

Finally, self-fulfilling beliefs cannot be analyzed merely in terms of indicative or

counter-factual conditionals. It is not enough for a belief to be self-fulfilling that, for a

certain subject in a certain context, were S to believe that p, p would be so. Such a

formulation is insufficient because is neutral about the order of dependence which explains

why the counterfactual holds. The counterfactual conditional 'if I were to believe that p, then

p' may hold because, since I am sensitive to the truth in the relevant domain, the only nearby

possible worlds in which I believe p are those in which I have successfully tracked the truth

of p. In such a case, the counterfactual 'if I were to believe that p, then p' holds in virtue of

the dependence of my belief upon the truth of what I believe. For a belief to be self-

fulfilling, however, the counter-factual conditional must obtain because whether p is so

---

[2] The limit case will be that in which the context is restricted to a belief at a particular time (e.g. what I *now* believe) when the context may restrict the relation of being self-fulfilling to beliefs which issue from some particular deliberation, such that whether p depends not on what S had believed, or what S believes while deliberating, but on what S believes *as the result of that very deliberation*. Importantly, if there is any first time when a belief becomes self-fulfilling, there will be a possible deliberation where it will become self-fulfilling. If one thinks of beliefs as states produced by specific judgments, or the conclusions of deliberations, the context relata might also relate the property of being self-fulfilling to those judgments, or conclusions. (There may be contexts in which, for example, I will be a millionaire just in case I *now* judge, or *now* deliberate and conclude, that I will be a millionaire; such cases will be important in the chapters which follow.)

depends on whether I believe that p, not the other way around. It must be that p would be so *in virtue of* S believing that p.[3]

Three more important features of self-fulfilling belief fall out of this formal characterization. First, we might have understood self-fulfilling beliefs as coming in a matter of degree, with some beliefs being 'partially' self-fulfilling, insofar as the fact that S believes that p at C make it more likely, or probable, that p. Believing that one will recover from one's illness, for example, may often make it more likely that a patient will recover from her illness, though it may only be sufficient to make it the case that she recover from her illness in cases in which the patient is otherwise sufficiently healthy.[4] In our formulation, however, the property of being self-fulfilling is reserved for those beliefs which, when held, not only increase the likelihood of their object, but actually suffice to make their object true.[5] The difference will often be a matter of the context relata. For a tennis player who is only barely outmatched, her belief that she will win might be able to make it the case that she will win, and so count as self-fulfilling, though in other contexts her belief, though it will make her perform better, will not be sufficient to make her win, and so not count as a self-fulfilling belief.

Second, the fact that p is so in virtue of the fact that S believes that p also means that often, though not always, self-fulfilling beliefs will involve a bi-conditional relationship between p and S's belief that p, such that p if and only if S believes that p. This is not always

---

[3] I understand in-virtue of here minimally, as neutral among a variety of asymmetric dependence relationships. (e.g. grounding, causation, constitution, etc.) I'll talk more about this in section III.D and in section II.

[4] As such examples show, this larger class is particularly salient if one is concerned with pragmatic reasons for belief -- those reasons for believing one has in virtue of facts which bear on whether having the belief would be good, or adventitious. Whatever pragmatic reasons one has for self-fulfilling beliefs, one will also have (if in slightly weaker form) for partially self-fulfilling beliefs.

[5] This is contra Johnston (1988) (and sometimes Reisner (forthcoming) and James (1897)) who, in their discussions of self-fulfilling belief, consider this larger class.

so, because even when p obtains in virtue of S's believing that p, p might be over-determined. It may be that your believing that you will win the tennis match would suffice for you to win the tennis match, even without your practicing, but that your practice would have also sufficed for you to win the match, whether or not you believed that you would. Or it may be that an eccentric billionaire will give you a million dollars if you believe that you will be a millionaire, but that your business acumen would make it the case that you would also be a millionaire even if it were not the case that you believed that you will be a millionaire. Or it may be that in believing that you are thinking, you will make it the case that you are thinking, though in doubting that you are thinking, you would have also made it the case that you were thinking.

Though I will occasionally consider such examples when salient to the particular issue at hand, in the following chapters I will generally assume that there is no such over-determination. I will assume contexts in which believing that p will be not only sufficient, but necessary, for p. I will do so because I hope to analyze what reason, if any, we have to adopt or retain a belief from the mere fact that the belief is self-fulfilling. Focusing the discussion on cases without over-determination lets us control for cases of self-fulfilling beliefs which are permissible, but for some independent or unrelated reason.

Third, it will often follow from the fact that a belief that p is self-fulfilling, both that were you to believe that p, then p, and also that were you not to believe that p, then not-p.[6] Self-fulfilling beliefs will thus satisfy the sorts of modal anti-luck conditions often associated with knowledge. Self-fulfilling beliefs will enjoy both *safety* and *sensitivity*: in all nearby possible worlds where you believe that p, p is true, and in all nearby worlds where p is false,

---

[6] Though not always. The mediocre tennis player may win the match in virtue of the fact that she believes that she will, though in most nearby worlds she loses, despite the contribution of her believing. (thanks to Tyler Burge for this point.)

you fail to believe that p.[7] In ordinary knowledge cases, of course, your beliefs are safe and sensitive because you are the sort of believer who discriminately tracks the truth. In the case of self-fulfilling beliefs, on the other hand, it is the other way around.[8] To be self-fulfilling is thus a more fine-grained property than the sorts of modal-properties with which some philosophers have tried to analyze knowledge.

§II. The Possibility of Self-Fulfilling Beliefs

One might object to such a formulation that there are not, indeed cannot, be self-fulfilling beliefs so described, because beliefs, to be *beliefs*, must be dependent on the part of the world which they are about, and not the other way around. In this section, I will canvas a series of objections along this line. I will argue that these objections rely upon a conflation of the various senses in which a belief might 'depend' on its object. Once we have clear the relevant senses of 'dependence' at issue, we can see that there is no trouble, in principle, with the possibility of self-fulfilling beliefs.

One way such an objection might be lodged is that self-fulfilling beliefs appear to lack a certain *objectivity* requisite for a propositional attitude to count as belief, properly speaking. Unlike matters of taste or preference, beliefs are objective in that they purport to express some fact of the matter in the world, and so can be evaluated as right or wrong independently of the agent. But if, as in the case of self-fulfilling beliefs, the truth of what the agent believes is a matter of whether the agent believes it, then the evaluation of the agent's belief as right or wrong *will* depend upon facts about the belief and the believer. The attitude will thus fail to admit the possibility of mistake in a way which may incompatible with an attitude's purporting to objectively represent the world.

---

[7] Nozick (1981). See also Duncan Pritchard (2005) for a more recent history.

[8] Fore a more in-depth discussion of this difference and it's import, see ch. 3.

However, the way in which a belief must admit of the possibility of mistake requires specifying. For one thing, the fact that a belief that p is self-fulfilling for some believer does not, at least *prima facie*, rule out the possibility of that believer being *unjustified* in believing that p, and in this sense believing in error.[9] For she may not know that her belief is self-fulfilling, and may otherwise have poor evidence for believing that p. Self-fulfilling beliefs fail to admit of the possibility of error, only if the error is one of truth and falsehood, not justification.

But the objection also cannot simply be that if a belief in some proposition, p, is self-fulfilling for some agent, then it is impossible for that agent to believe that p and be incorrect, in the sense of believing falsely. For beliefs about necessary truths, like the truths of mathematics, will be such that it will be impossible, in believing them, for *any* agent to believe them and be in error. This, however, casts no doubt on the objectivity of mathematical beliefs.

The distinctive problem with a self-fulfilling belief that p is not that the belief could not have been incorrect, but that the *believer* could not have failed to be correct. That is, the problem is not that if the subject believes that p, that propositional attitude couldn't have been mistaken, but rather that no attitude the agent might have had with respect to p could have been mistaken. The agent would have had a correct doxastic attitude with respect to p, whether she believed that p or disbelieved that p. In this self-fulfilling beliefs differ starkly from necessary truths. You cannot err in believing 2 + 2 = 4, but of course you will thereby err in disbelieving 2 + 2 = 4.

The fact that in the case of self-fulfilling belief you will be correct no matter your attitude may seem to challenge the idea that self-fulfilling beliefs really are properly

---

[9] At least, on any but the most extreme externalist view. Though more nuanced views may hold that, though the justification is non-trivial, the fact that a belief is self-fulfilling will in some more complicated way ensure that the believer is justified.

understood as beliefs. Beliefs are correct, at least in part, because their propositional object is true. Whether a proposition is true is a matter of whether it corresponds to how the world is, not the attitudes of the believer. And since the world is mind-independent, it may appear that it must be possible for the agent's attitude to be mistaken, if it is really purporting to represent some truth about the world.

But this objection fails to mark an important difference between the claim that what it is for some belief to be correct is a matter of whether an agents believes it, and the weaker claim that *what makes* the believed proposition true (and so makes a belief correct) is that the agent believes it. The first of these is a conceptual claim about nature of correctness incompatible with the objectivity of belief, the second is a much less exceptional claim about the conditions which satisfy the independently specified truth-condition for correctness.

The issue can be cleared by keeping straight the difference between the belief (an attitude or psychological state), the believed proposition, and the state of affairs represented by that proposition. Beliefs are correct if their propositional object is true. And whether that proposition is true a matter of whether the state of affairs the proposition represents obtains. And so a belief is objective in the sense that whether or not a belief is correct is a matter of whether or not a certain state of affairs in the world obtains.

It is compatible both that the criteria of correctness of an attitude be specified independently of the belief or the believer (e.g., in terms of truth of its propositional object), and yet, because of the structure of the world, that the belief or believer effect whether some state of affairs obtains, and so effect whether those criteria are met. In such a case, the fact that the criteria for correctness are specified independently of the belief allows us to understand the possibility of the believer to be in error, even if the structure of the world,

specifically the connections between the state of affairs of S believing p and the state of affairs of p, guarantees no such error will occur.

This point can be seen more clearly by dropping the misleading temporal metaphor when speaking of the self-fulfilling belief 'becoming' true when believed. Barring some special thesis about the openness of future contingents, I assume the truth of a proposition to be eternal. When a belief that p is self-fulfilling, the proposition, p, will be true long before S ever believes that p, and will remain true long after. It is just that p will be eternally true *because* S believes that p at the appropriate time, making the represented state of affairs obtain.

A related but more sophisticated objection that might be lodged against self-fulfilling beliefs involves considerations of direction of fit. Beliefs (and cognitive attitudes more generally) are representations which have a *world-to-mind* direction of fit. They are representations which are supposed to correspond to the way the world is. In contrast, desires (and conative attitudes more generally) have a *mind-to-world* direction of fit.[10] They are representations with which the world is supposed to correspond. Self-fulfilling beliefs may appear to be on the wrong side of this divide. Since, for self-fulfilling beliefs, how the world goes depends on how we represent it as going, rather than the other way around, they may seem not of the right kind to be beliefs at all.

But, similar to the first objection, the problem with this new objection is that it fails to disambiguate two ways in which the world and our representations might depend on one another. The first way involves the order of *existential* dependence. Our representations may bring about (by causing, or generating, or grounding, or constituting) the existence of the relevant part of the world, or vice-versa. My desire that it be warmer, for example, might

---

[10] For an overview, see Humberstone (1992).

cause me to go to sunnier climates, whereas my belief that it is warmer might result from my perception of sunnier climates. The second way involves the order of correctness. The way the world is might explain whether our representation is right or wrong, or vice-versa. An example of the latter seems to be what Anscombe has in mind in her famous passage in *Intention* where she discusses the difference between the man going around town with a shopping list in hand buying items, and the detective following him around listing the items he buys. The difference, she says:

> is precisely this: if the list and the things that the man actually buys do not agree, and if this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance…whereas if the detective's record and what the man actually buys do not agree then the mistake is in the record.[11]

For Anscombe, both intentions and beliefs are such that, when the world and one's representations fail to match up, a mistake is involved. The difference between the attitudes lies in where the mistake is to be located. In the Detective/belief case, the mistake was to represent the man as purchasing margarine when in fact he purchased butter. In the Shopper/intention case, the representation was a good one, and the mistake was in the shopper's failing to purchase butter, as he had represented himself as doing. (As Anscombe observes: "if his wife were to say: 'look, it says butter and you have bought margarine', he would hardly reply: 'what a mistake! We must put that right' and alter the words on the list to 'margarine.'")

What is crucial to belief, I think, is this second order of dependence. Belief aims at the truth in the sense that whether or not beliefs are correct is a matter of whether or not they correspond with the world. It is essential to beliefs, as beliefs, that they purport to represent the world, not that they be *generated by,* or result from, the part of the world they

---

[11] Anscombe (1963)

purport to represent. And it may be that the order of existential dependence goes one way, when the order of correctness goes the other.[12] A belief may aim at the truth, and so depend on whether the world is a certain way for its evaluation as correct or mistaken, even though how the world is turns out to depend on the belief.

One might accept that beliefs need only depend on the world for correctness, not existence, but remain skeptical of self-fulfilling beliefs because of a skepticism that these two directions of dependence really can come apart completely.[13] In order for a state to aim at the truth, in the sense of being correct if it represents what is so, one might think, the state must do so by functioning to bring the representations into correspondence with some antecedent way the world already is.[14]

This skepticism, I think, gets its strength from taking too seriously the metaphor of 'aiming.' The strict sense in which belief aims at the truth is the sense in which what it is for a belief to be true, and so correct, is for the content of the belief to be in agreement, or accordance, or correspondence, with the world. The skeptic has added to the original picture by assuming that there must be some target out there, independent of the belief, for the belief to track, reflect, or aim at, so that it is the belief which must be brought into

---

[12] Here, as in much of the present work, the point is anticipated by David Velleman. Velleman (2000) makes a roughly similar distinction between the order of existence and order of correctness, in terms in terms of "direction of guidance" versus "direction of fit," respectively. (pp 24-26) I address the particulars of Velleman's views in chapter 3.

[13] And as many do think. See, for instance, Jonathan Adler (1999)

[14] I say 'be the kind of attitude' because the skeptic cannot think that *every* representation, to be belief, must be caused by the world. Notoriously, not all beliefs exist as a result of the way the world is. Some beliefs are false, and are false because they've failed to track the world. Some hallucinatory beliefs appear unconnected to the world entirely. But self-fulfilling beliefs are not like this. They are not unconnected to the world. Far from it, the way they represent the world will often correspond tightly to the way the world is (as their modal properties attest). The same is true of successful desire representations. The problem, if any, is that self-fulfilling beliefs, like successful desires, and unlike paradigmatic perceptual beliefs, lead the world, rather than the other way around.

correspondence with the world, and not the other way around.[15] But this last step involves a substantive addition to the notion that a belief that p is correct just in case p is true.

This addition may seem to follow straightforwardly from a realist picture of truth. For realism is often glossed as the position that there is an objective, mind-independent world out there to track. But as we have already seen, this gloss is not strictly correct. We are not Cartesian egos, apart from the world and observing it. We ourselves are also a part of the world we are trying to represent, as are our beliefs. For that part of the world made up of our beliefs, and the parts of the world effected by our beliefs, what we believe *will* effect what the objective world is like. Sometimes our belief about the world, though aiming to get at the truth, doesn't simply track the truth; it creates it. When how we think so effects the object of that belief, the resulting beliefs may be self-fulfilling – they will be true because they are believed. This does not mean they are any less beliefs – representations which purport to represent reality as it is, and which we judge according to their success in doing so.

§III. The Variety of Self-fulfilling Belief

Assuming, then, that that there can be such a thing as a self-fulfilling belief, we can now inquire into whether there are any. How can our beliefs effect the parts of the world we are believing about, and what kinds of beliefs might effect the world in such a way so as to make themselves true. I will start by considering the sorts of cases where beliefs are sometimes claimed to be self-fulfilling, and then see what generalizations we can draw.

---

[15]this moral-psychological claim: that beliefs which don't function to track the world cannot be beliefs, needs to be distinguished from a related epistemological claim: that such beliefs, while possible, could never be justified, because for a belief to be *justified* there must be something about the world, tracked by the believer, which made the believer come to the belief. I'll save discussion of this second claim for later chapters concerning the epistemic status of self-fulfilling beliefs, but it is worth bearing them in mind together, as I suspect that the epistemic claim draws much of its prima facie plausibility from an implicit assumption of the moral-psychological claim.

14

A. Self-Knowledge

One place we can find attributions of self-fulfilling belief is within discussions of self-knowledge of our beliefs or mental states more generally. The claim that such beliefs are self-fulfilling is most often made with regard to cogito-style beliefs; beliefs like the belief that I am thinking, or the self-referential belief that I believe this (very belief). Tyler Burge, for example, claims that in such cases "once one makes the judgment... one makes it true" where this holds in virtue of a "constituitive relationship between judgments and their subject matter -- or between the judgments and their being true."[16]

In these cases, the belief makes its content true because the object of the agent's belief constitutively depends on the agent's belief. When an agent believes that she is thinking, for instance, her believing constitutes her thinking, and so makes it the case that her belief is true. For Cogito-style beliefs more generally, the agent's believing that p is, (at least in part) *what it is* for p to be so. And so the fact that the agent has the belief will, in virtue of the conceptual connections between the belief and the content of the belief, make the belief true.

The claim that one's beliefs are self-fulfilling is also attributed to cases of one's present tense beliefs about one's own mental states more broadly. Such a position has been recently advanced by Richard Moran and Matthew Boyle, who advance the claim that, more generally, "self-consciousness has special consequence for the object of consciousness."[17] On at least one interpretation of these theories, when we form second order beliefs about

---

[16] Burge (1996) 92;98. See also Burge (1988). Burge refers to the property of being "contextually self-verifying" rather than "Self-fulfilling" though, for Burge, a judgment can be contextually self-verifying not only when one judges but when one "just engages in the thought." The property of being contextually self-verifying thus applies to cases in which the truth of p is over-determined, like the belief that I am thinking, as well as cases where it is self-fulfilling in the more restricted sense laid out section 1, like the reflexive case where I believe that I believe this (very belief).

[17] Moran (2001). Pg 28. Boyle (2011).

our own beliefs, in the appropriate circumstances, the first order judgment is a component of the second order judgment, so that when we believe that we believe that p, in doing so we 'make up our mind' about p, and thereby make it the case that we believe that p.

B. *Practical Knowledge*

A similar claim to be self-fulfilling is sometimes made about cases of practical knowledge of our actions action; beliefs of the form: I will φ, or I am φing, where φ is some possible action or activity. One prominent traditions holds that we have knowledge in these cases because our beliefs are the "cause of that which they understand," that is, because these beliefs are self-fulfilling.[18]

These beliefs may be self-fulfilling because, on a number of theories of action, whether we φ is a matter of whether we believe that we will φ, in virtue of some psychological connection between our beliefs and our actions. On one of the most thorough such accounts, David Velleman's, this psychological connection is realized through the agent's desire to know what they are doing.[19] Given this desire, an agent's belief that they will act a certain way will motivate an agent to be or act as they believe themselves to be. In normal circumstances, it will interact with the agent's other desires, re-enforcing some and constraining others, in such a way as to ensure that the agent acts as they believe that they will, because they believe that they will. Thus in believing that one will φ one will, in ordinary circumstances, make it the case that one φs.

---

[18] See for instance Aquinas (1247), Anscombe (1967), Hampshire (1959), Harman (1976), and Velleman (1989). Apropos of the previous section, it is not clear that everyone in this list thinks of the representations which are "the cause of that which they understand" as *beliefs*, though at least the last two, Harman and Velleman, do so explicitly.

[19] Velleman (1989)

The possibility of these sorts of psychological connections between our beliefs and other parts of our mental life also makes possible a number of related self-fulfilling beliefs. "Positive Thinking" effects, in which an agent performs better because she believes she will, will hold in virtue of some sort of psychological laws. Presumably some placebo effects, in which a patient's belief that she will recover helps make it the case that she recover, are efficacious through some psychological regularity as well.[20]

A slightly different, but related set of cases, involve self-fulfilling beliefs which make themselves true through our beliefs' interactions with other peoples' psychologies. Since other people can recognize and respond to our own beliefs in regular ways, if they respond in these ways to our belief about how they will respond, these beliefs will be true in virtue of our having them. Thus, our beliefs about other people can sometimes be self-fulfilling, just like our beliefs about ourselves and the parts of the world under our active control can be. Many such cases will be quotidian. A child's belief that she will be fed may prompt a parent to begin her feeding. But this general phenomenon is also what explains a number of cases in the literature. It is behind, for instance, William James' confident agent, whose belief that she will be liked may be self-fulfilling, if those the agent interacts with recognize and respond to the confident belief by liking the agent.[21] It may also explain cases of self-fulfilling belief in the psychology of education, in which student's performances match the expectations of their instructors.[22]

---

[20] See, for example, Scheier, Matthews, et. al (1989)

[21] James (1897)

[22] See, for example, Wilkins (1967)

Many psychological cases of self-fulfilling belief, like examples involving beliefs about the success of a marriage,[23] may be mixed; they will be self-fulfilling because of how both you and others respond to the belief. I will not attempt to analyze these joint cases here, though such cases can be understood relative to one single subject, if the attitudes of the other agents are made part of the context relata.

C. Science Fiction Cases

The final sort of case most often discussed are a subclass of the "Science Fiction" cases featured in so many philosophical thought experiments. We can consider the further exploits of Gregory Kavka's "eccentric billionaire" who, equipped with "the latest 'mind reading' brain scanner and computing device designed by doctor X" will, by giving you a million dollars if you believe that you will be a millionaire, ensure that your belief that you will be a millionaire will make it the case that you will be.[24] Or we can consider the neurological illness of Andrew Reisner's patient, "improbably… connected in particular to brain states that encode beliefs about this specific illness" such that patients who believe they will recover will, and those who fail to never will.[25]

What these science fiction cases have in common is that they all take advantage of the fact that our beliefs either reduce to, are realized by, or in some way involve some physical correlates. Since those physical correlates will be causally efficacious in the normal sorts of ways, there is the possibility that the physical correlate can causally interact with whatever the object of the belief is in ways which make the belief true. This is how my beliefs about my future finances sets off the billionaire's scanner, so that he can give me the

---

[23] See Moller (2003) or Merusic (2013)

[24] Kavka (1983); See Johnson (1988) for an example of such an amended case.

[25] Reisner (forthcoming)

million dollars, and this is how my beliefs about my health, through the associated neural activity, make it the case that I am cured of my disease.

If then, we imagine sufficiently circuitous and unlikely causal chains, we can leverage the brute causality of beliefs (or their associated physical correlates) in such a way as to imagine that a belief in almost any proposition might, in the appropriate context, for the appropriate subject, be self-fulfilling.[26]

D. Common Ground

There are some common trends in the various cases surveyed above. First, the most likely self-fulfilling beliefs tend to beliefs about the subject, or matters closely related to the subject. This is no co-incidence, since the most common way for a belief to effect the world is through the believer.

Second, self-fulfilling beliefs are also beliefs which are common candidates for attributions of some special epistemic status. We have already mentioned that self-fulfilling beliefs will be a sub-class of beliefs which enjoy the epistemic modal properties of safety and sensitivity. But cases of self-knowledge, and practical knowledge, are also cases where an agent is often thought to enjoy a special sort of "first-personal authority" over their beliefs, as well as a certain "immediate" or "non-observational" knowledge of the subject matter. This special epistemic status of self-fulfilling beliefs will the subject of the chapters to come.

Still, as these cases illustrate, a variety of kinds of beliefs can be self-fulfilling, and these self-fulfilling beliefs can differ in important ways. The connection between the belief and the world, through which the belief makes itself true, are sometimes psychological,

---

[26] Such cases need not even involve material causation, if you accept sufficiently outlandish scenarios. Instead of the Scientist with the brain-reading machine, we could equally well suppose an occasionalist story, on which some Demon or Angel decides to make it the case that whatever you believe is so. In such circumstances, all of your beliefs may very well be self-fulfilling.

sometimes conceptual, and sometimes causal. Accordingly, the connections can vary widely in character. Sometimes the connection between the belief and the believed state of affairs is contingent, as in the psychological cases; sometimes it is necessary, as in the conceptual connections between cogito-style beliefs and their contents. Sometimes the connection is nomonologial, sometimes, as with the science fiction cases, the connection is brute. Sometimes we are intimately involved in making the belief true, either as agents (as in the practical knowledge case), or as patients (as in the placebo cases). Sometimes the connection is more indirect, holding through external agents (as in science fiction cases).

What underlies each of these otherwise disparate cases is the distinctive order of dependence between the belief and the world. Regardless of mechanism, the result, if a belief is self-fulfilling, will always be that there is no fact out there to be believed, independent of the agent's belief.

It is my hunch that this is a philosophically important division. I suspect that these beliefs, and especially their epistemic status, are best understood not in terms of the nature of their connection with the world, or their subject matter, or their modal properties, but in terms of their being self-fulfilling. In the proceeding chapters, I hope to show that the variety of self-fulfilling beliefs are amenable to common analysis, and that thinking of such beliefs *as* self-fulfilling will bear fruitful results for understanding their epistemic status.

§IV. Further Investigation

Before going on, I hope to say a word about how I will go about investigating self-fulfilling beliefs. The variety of different beliefs which might be self-fulfilling beliefs means that such an investigation faces a particular methodological challenge. As mentioned above, the most common way for a belief to be self-fulfilling is for it to be about the subject's mental life or actions. This means that the most central cases of self-fulfilling belief will be

20

philosophically contentious and the less controversial cases, like the science-fiction cases, will be peripheral, and may appear to lack philosophical importance. In the proceeding discussion, I have chosen to focus most of my discussion on the science fiction examples, returning to the more central cases of practical knowledge and self-knowledge only in the final chapters.

There are two reasons for this choice. First, because the eccentric billionaire case involves a belief for which we have obvious pragmatic incentives, it will allow for discussion of the relationship between epistemic and pragmatic reasons for belief. This may complicate the example, insofar as we are studying primarily the epistemic status of self-fulfilling beliefs, but it is a relationship which is central to much of the literature's discussion of the topic, and which will become increasingly important in later chapters.

Second, because of their artificial construction, the only special thing about the science fiction cases, unlike more central cases, seems to be the fact that they are self-fulfilling. This will allow us to isolate our subject-specific intuitions, and to study what reasons we have for adopting a belief from the mere fact that a belief is self-fulfilling. Since they are in this way the simplest case, what I say about the science fiction cases should follow *a fortiori* for the rest of the beliefs surveyed above, insofar as they too are self-fulfilling. If we show there is a special sort of epistemic status to these cases, we will have shown that this status, in more complicated classes of beliefs, might also be explained merely in virtue of their being self-fulfilling.

Chapter 2: Evidence and Self-Fulfilling Belief

Suppose an Eccentric Millionaire comes to your door with the following credible offer: she will give you one million dollars if you believe that you will be a millionaire.[1] Should you believe that you will be a millionaire? The answer to this question is complicated by the fact that the belief that you will be a millionaire is self-fulfilling. It belongs to the class of beliefs whose contents will be true just in case you believe them, because you believe them.[2] How, when belief in some proposition is self-fulfilling, ought we to engage in theoretical deliberation about, and so come to a belief about, that proposition?

§I. Preliminaries

Before delving into these issues more deeply, I will first try to narrow the discussion's scope by setting aside two sets of issues with which I will *not* be concerned. One set of issues with which I will not concern myself directly involves pragmatic reasons for belief. Traditionally, what one thinks of this kind of millionaire puzzle comes down to what one thinks about at least one tempting reason to believe that you will be a millionaire: that being a millionaire would make you much better off than not being a millionaire. Pragmatists, most famously William James, have been sympathetic to the force of this reason. James argues that, at least in cases of self-fulfilling belief, one is permitted to make a qualified 'leap of faith,' believing whatever is in line with one's needs.[3]

Many have remained unpersuaded. This is, at least in part, because this kind of pragmatic reason is in apparent conflict with a standard account of epistemic rationality on which reasons for belief must involve evidence for the believed proposition. I am not going

---

[1] An example from Mark Johnston (1988), 68. A number of philosophers have presented puzzles in a similar vein. See, for example, Anscombe (1957), Velleman (1989), and Reisner (2013).

[2] I characterize such beliefs more fully in chapter 1.

[3] James (1895), 57.

to involve myself in questions about the possibility and force of pragmatic reasons here. I am going to restrict myself to asking what we ought, epistemically speaking, to believe with respect to the proposition that we will be a millionaire.[4]

A second set of issues I wish to set aside involves certain aspects of the epistemic status of self-fulfilling beliefs with which I will not concern myself. The millionaire example involves an agent actively engaged in explicit theoretical reasoning about whether she will be a millionaire. I want to ask how an agent ought to engage in such reasoning. This means that the epistemological inquiry will be narrow in certain ways.

First, many beliefs do not involve explicit theoretical deliberation, but are instead the product of some other lower-level truth-directed process. While much of the proceeding discussion of the rationality of theoretical deliberation will extend to discussions of the warrant of such lower-level belief forming processes, I will not be concerned with that issue here.

Second, my inquiry will exclude a variety of questions about some of our wider epistemic obligations. There are interesting questions about when we are obliged to gather evidence and questions about when and whether we are obliged to consider the truth of a proposition (especially arcane or unimportant propositions) and so obliged to believe them. In this case, the agent is already actively engaged in theoretical deliberation about the truth of some proposition, and so these questions, while interesting, will not apply.[5]

---

[4]This question of what we ought, epistemically speaking, to believe with respect to self-fulfilling beliefs is of course deeply relevant to the pragmatist-evidentialist debate. Prior to evaluating the pragmatist-evidentialist debate, we need to see how much disagreement there is, which means we need to know what attitude the evidence dictate we adopt. One upshot of this chapter will be that there may be much less straightforward disagreement between the evidentialist and pragmatist than many have thought.

[5] For discussion of these issue, see Feldman (1988) and Hall and Johnson (1998)

Finally, and most importantly, the puzzle we are considering is formulated in the first-person. I am interested in what I (or we) ought to believe, given that whether or not we believe that we will be a millionaire will determine whether we will be. The investigation is in this way explicitly internalist in nature. It is concerned, from within the perspective of the agent engaged in the theoretical deliberation, with the sorts of reasons she might employ in reasoning about whether p is so, and the sorts of norms which might guide her in that reasoning. In particular, what are we to say about what our evidence dictates we conclude in such a case?[6]

In the following sections, I will explore this question in more detail. I will argue that in the case of self-fulfilling beliefs our evidence fails to dictate any attitude at all. Our evidence is insufficient to support belief, disbelief, or even suspended judgment, and so our evidential reasons under-determine what attitude we are to have. This exploration will thus suggest the need for a partial re-thinking of the standard picture of epistemic rationality: in the case of self-fulfilling belief, if our reasoning is to aim at the truth, it must do so by going beyond the reasons of evidence.

My argument will proceed in three parts. In the first part, I will develop and consider an increasingly popular line of reasoning which holds that self-fulfilling beliefs are epistemically impermissible because adopting them will involve an illicit form of bootstrapping.[7] Though your belief that you will be a millionaire will be true if held, in order for you to take that conditional as evidence for believing that you will be a millionaire you would need to have the belief already, and so cannot believe it for that reason.

---

[6] A distinction which has sometimes been understood in terms of differences between regulative/evaluative norms (see e.g. Goldman (1980)).

[7] See, e.g. Grice (1971); Johnston (1988); Foley (1991); and Reisner (2013).

In the second part, I will trace this objection to a particular model of theoretical deliberation: a model which pictures deliberation as a function, treating the deliberation's inputs as given, fixed prior to and independently from the deliberation. Though such a picture may seem plausible, I will argue that self-fulfilling beliefs can help us see how such a function-model is inadequate for a general understanding of theoretical deliberation. Because it is unable to accommodate the possibility that the outcomes of our deliberation can sometimes be evidence for the proposition we are deliberation about, the function model is unable to account for cases where what we believe affects the truth of what is believed.

Finally, I will consider how we might reason about the millionaire case if we were to move beyond the function model. I will suggest that with a better picture of theoretical deliberation, we can see that for self-fulfilling beliefs, our evidence does not dictate disbelief, but rather under-determines any appropriate doxastic attitude. Our evidence is insufficient to permit belief, disbelief, or even suspended judgment.

§II. Evidence and Bootstrapping

Usually, believing some proposition because someone will give you a million dollars to do so will be straightforwardly incompatible with believing on the basis of your evidence. Suppose, for instance, that you were offered one million dollars to believe that pigs can fly. If you believe that pigs can fly because you will receive a million dollars to do so, you will be holding a belief which goes against all of your (presumably substantial) evidence to the contrary.[8]

But having the belief that you will be a millionaire on the the grounds that the you will receive a million dollars to do so will not obviously be in conflict with believing on your

---

[8] It is not at all clear that such a feat is even possible. See, for instance, Kavka (1983)'s discussion of the toxin puzzle, and more recently Hieronymi (2005)'s discussion of believing at will. But even if such a feat is possible, it is certainly not epistemically permissible.

evidence in the same way. The fact that you will receive a million dollars to believe you will be a millionaire can show that it is *true* that you will be a millionaire, if you believe that you will be, and so count as evidence for your belief that you will be a millionaire.

Self-fulfilling beliefs may thus appear to be in particularly good epistemic shape. Because a belief that p is self-fulfilling, p will be true whenever it is believed. And so if an agent knows that her belief that p is self-fulfilling, she will have good evidence for her belief.

Recall, however, that we are concerned with the question of what belief the evidence dictates, from within the perspective of the agent engaged in the theoretical deliberation. We are concerned with evidence, insofar as it connects to some regulative epistemic norm, meant to regulate the agent's reasoning about whether p is so, and so meant not just to govern the permissibility of an agent holding some belief, but of her *adopting* a belief. When we ask what the evidence dictates, from the agent's point of view, we are not asking, of some already existing belief, whether it is supported by the agent's evidence. Rather, we are asking, of an agent, whether she has sufficient evidence available to employ in forming that belief in the first place.

Given this concern, the relationship between self-fulfilling beliefs and one's evidence begins to appear more problematic. For while it may be true that once you hold a self-fulfilling belief, you will thereby have good evidence to support your belief, you do not have this evidence yet, when deliberating, because you do not yet have the belief that you will be a millionaire. And so it is unclear whether you have sufficient evidence available now to adopt the belief in the first place.

It is unclear, in part, because while the issue is often understood in terms of *ex-post/ex-ante* or *synchronic/diachronic* norms, such terms map onto the present issue imperfectly. This is because the norms regulating permissible belief adoption and the norms by which we

might evaluate the permissibility of holding some existing belief might each be understood in terms of one another. We might evaluate an existing belief in terms of how it was formed, and we might articulate the diachronic norms regulating permissible belief formation in terms of whether the resulting beliefs *will be* supported by evidence, once held.

Still, if we are concerned with the perspective of the reasoning agent, we can see that in cases of self-fulfilling belief it could not be enough for the agent to adopt the attitude toward p which will be supported by evidence, once held.

For in the case of self-fulfilling belief, there will be no single attitude the agent might adopt which will be supported by evidence, once held. If the fact that an agent's future belief that she will be a millionaire would be supported by evidence gives an agent a reason to believe that she will be a millionaire, that agent will have the very same reason to believe that she will not be a millionaire. For if she believes that she will not be a millionaire, she won't be. And once she believes that she will not be a millionaire, the fact that the belief is self-fulfilling will be evidence to support her belief that she won't be a millionaire.

More generally, in cases of self-fulfilling belief, where the truth of a proposition depends on whether an agent believes that proposition, two contrary doxastic attitudes the agent could hold with respect to p might each be correct, and supported by evidence, once they were held. Thus, if an agent is permitted to adopt any attitude which will be supported by the evidence, then if a belief is self-fulfilling, our evidence cannot be sufficient, on its own, to dictate for an agent to whether to believe or disbelief that p. Facts about the evidence alone would under-determine what to believe with respect to self-fulfilling beliefs.[9]

---

[9] Several philosophers have been tempted to accept that our evidence underdetermines which doxastic attitude a rational agent is to adopt, but deny that this underdetermination poses an important problem. (See e.g. Velleman (1989); Reisner (2013); Wedgwood (2013); Peels (2014).I discuss such responses in the next chapter.

Suppose instead that we understand permissible belief adoption not in terms of what evidence one will have to support the belief once it is held, but in terms of the evidence one already has available in support of p when one begins reasoning about whether p is so.

Self-fulfilling beliefs may now appear to be in rather bad epistemic shape. For while you will have evidence that you will be a millionaire once you believe it, you lack any such evidence that you will be a millionaire now. This means that it cannot be permissible to adopt the belief that you will be a millionaire: acquiring the belief would require an illicit bootstrapping. Grice makes such an objection forcefully against self-fulfilling beliefs in "Intention and Uncertainty." Imagining a case in which my going to London depends on my believing that I will go, Grice observes:

> If my going to London is to depend causally on my acceptance that I shall go, the possession of satisfactory evidence that I shall go will involve possession of the information that I accept that I shall go. Obviously, then, I cannot (though others can) come to accept that I shall go on the basis of satisfactory evidence; for to have such evidence I should have *already* to have accepted that I shall go.[10]

The problem is that, while it may be true that once you hold a self-fulfilling belief, you will thereby have good reasons for your belief, that will be the *only time* that a belief's being self-fulfilling will give you good reason for the belief, and so you will never have reason to adopt the belief in the first place.

Consider the point at which you first entertain the question of whether you will be a millionaire, after being given the millionaire's offer. In this case, you plausibly do not yet have the belief that you will be a millionaire, which means it is likely not the case that you will be a millionaire. It is true that, if you had the belief, you would then have it for good reason. Moreover, since you know the belief is self-fulfilling, it's true that you already have

---

[10] Grice (1971), 274. For a similar assessment, see also Foley (1991), Johnston (1988), and Reisner (2013). Interestingly, while Reisner accepts that, in cases like these, our evidence dictates disbelief, he does not ultimately agree that in such cases a rational agent constrained by her evidence, arguing instead that such cases permit a "limited version of doxastic voluntarism." I address this position in chapter 3.

good reason for believing the conditional: if I believe that I will be a millionaire, then I will be a millionaire. But none of this will be sufficient for you to permissibly adopt the belief that you will be a millionaire because, although you believe the conditional, you do not yet believe the antecedent. Since the antecedent in this case just is the belief you are trying to acquire, you cannot permissibly use this conditional to get to the conclusion that you will be a millionaire, because you would have already needed to have believed the conclusion (or at least, have had good reason to believe that you believed the conclusion) in order to believe the antecedent. You would need to pull yourself up by your own bootstraps.

It might look as though this bootstrapping argument would show that a belief's being self-fulfilling has no bearing on whether to adopt the belief. If your evidence had counted against p, you should keep on disbelieving that p, and if you lacked any prior evidence with respect to p, or if your evidence was counter-balanced, you should keep on suspending judgment with respect to p.

In fact, for most self-fulfilling beliefs, it shows something stronger. For if you know a belief that p is self-fulfilling, there will be reason to conclude that p is so if you believe that p, but reason to conclude that p is false in any case where you fail to believe it, (whether you fail to believe it in virtue of disbelieving p, suspending judgment with respect to p, or lacking any doxastic attitude at all toward p).[11] So if, as in the present case, you know that you do not yet have the belief that p, learning that the answer to the question of whether p depends on whether you believe that p should not only fail to give you reason to adopt the belief that p,

---

[11] I assume here there is no over-determination (perhaps, believing that I will become a millionaire would make it the case that I would be, but my financial acuity ensures that I will be regardless) so believing that p will be not only sufficient, but necessary, for p. I make this assumption not because of its plausibility for my expected academic audience, but to control for cases of self-fulfilling beliefs which are permissible, but for some independent or unrelated reason.

it should provide you with a reason to conclude that p is *false*.[12] Even if you go into the deliberation merely not believing that you will be a millionaire, you should go out believing that you will not be a millionaire.

This is a result which would have strong and surprising consequences for the debate between the pragmatist and the evidentialist. It suggests that a positive answer to the millionaire case would commit pragmatists to the claim that you may believe not just beyond, but *against* your evidence: believing that p when the evidence not only fails to support your belief, but supports believing not-p instead. This surprise would also be bad news for the pragmatists. For while pragmatists disagree that a belief is permissible only if it is supported by one's evidence, most do not wish to deny the weaker evidentialist principle that one should believe that p if p is supported by one's evidence.

We have thus arrived at a provisional conclusion to our investigation. When a belief that p is self-fulfilling, the evidence dictates that an agent disbelieve that p, for any other attitude would require an illicit form of bootstrapping. In what follows, I will argue that this bootstrapping objection is importantly mistaken. Understanding why can give us an insight into a more fundamental problem with the way many have been thinking about theoretical deliberation, and of the norms of theoretical rationality.

§III. Theoretical Deliberation

The problem with bootstrapping objection is that it presupposes a very particular view of the theoretical deliberation which underlies belief adoption. And this view, while attractive, is mistaken.

I will call the mistaken view the *Function Model* of deliberation. The picture I have in mind is of the kind which finds expression in Goldman (1980):

---

[12] A consequence noted by Reisner (2013).

the regulative justificational status of a doxastic attitude for person S at time t depends upon (a) the right set of doxastic instructions and (b) the states S is in at (or just before) t…we may represent [a principle of justification] as a function whose *inputs* are certain conditions of a cognizer – e.g. his beliefs, perceptual field, and ostensible memories – and whose *outputs* are prescriptions to adopt (or retain) this or that doxastic attitude – e.g. believing p, suspending judgment with respect to p, or having a particular subjective probability vis-à-vis p[13]

The Function Model treats an agent's starting beliefs as given, fixed prior to and independently from the deliberation. When we deliberate, we take all of our relevant beliefs, and see what belief to produce as a result. On such a view, the process of belief adoption and revision is like a function, from one set of beliefs to another, with the principles of justification telling you what outputs you should end up with.[14]

There are two important and related features of such a model that are worth mentioning, since, while they may seem initially innocuous, they will lead to serious doubts about the model's applicability to deliberations about self-fulfilling beliefs.

One, *the model treats deliberation as a series of discrete steps.* On a simple model, there are three parts to a deliberation: First one takes stock of one's reasons, "gathering" them as inputs. Then one employs those reasons and deliberates. Finally one finishes (if one deliberated correctly) by producing the appropriate doxastic state as an output.

Two, *the inputs are exogenous to the model.* On any version of the model, the inputs will be treated as givens, fixed prior to and independently from the deliberation itself. This will follow in part from the fact the model goes in steps. Since the inputs are gathered first, they will be the beliefs that one had at the start, when entering into the deliberation; they will not include beliefs you hold at some later time, including beliefs that you are to come to during or as a result of the deliberating.

---

[13] Goldman (1980), 27-29.

[14] For simplicity, I will restrict my discussion to beliefs, though I leave open, like Goldman, the possibility that a function model might also include in its inputs a broader range of representational states including, e.g., perceptual representations.

Together, these features ensure that the adoption of self-fulfilling beliefs would inevitably require an illicit bootstrapping. This is because successfully reasoning to a self-fulfilling belief would require two items: belief in the self-fulfilling conditional: "if I believe that p then p," and the second-order belief: "I believe that p." If this second-order belief is not among ones reasons, having the conditional 'if I believe that p then p' will never get you to the belief that p as an output. And having the conditional 'p only if I believe that p' will always give you a reason to disbelieve p. But, crucially, on the function model, this second order belief is precluded from our reasoning by the very nature of the model. If it is not entered in as an initial input, it cannot enter into the deliberation when we bring our reasons to bear. And because the inputs are exogenous, it would be impossible for some fact about what you will later believe, as a result of this deliberation, to be entered in as an initial input.

The problem is that the function-model assumes that we can always bring to bear all the relevant information we need to decide what to believe as exogenous inputs. While such a picture may work well in many normal cases where our deliberation is tracking some mind-independent truth, it is ill-equipped to handle beliefs whose truth depends on the outcome of the deliberation itself. For if the outcome of the deliberation is relevant to the truth of the object of deliberation, then in order to include all the relevant facts as *prior* inputs, a deliberator will be required to answer their question before they begin.

The bootstrapping objection gets off the ground because the function model precludes us from entering into the model what we know to be the one crucial consideration for determining whether we will be a millionaire: whether we now believe that that we will be a millionaire. But since the model supposes that every relevant consideration can be modeled as an input, the absence of the consideration: 'that I believe that p' appears to count as support for disbelieving p. But this is a glitch of the model. The belief that p is only

absent as an input because beliefs involving the outcome of the deliberation are inadmissible as a structural feature of the model.

But if a model of deliberation is unable to model all the relevant factors that might go into a theoretical deliberation, this shows not that it is impermissible to believe the proposition we are deliberating about, but that we have chosen a poor way of modeling deliberation. It is a problem, not for self-fulfilling beliefs, but for the function-model of deliberation.

Suppose that we abandon this artificial demarcation, and assume we *can* employ in our reasoning facts we arrive at while deliberating. What will happen? The result, I think, is an interesting one. Since we know that p is self-fulfilling, we need to determine what we believe about p, to determine whether p. We know we hadn't believed p thus far, but we need to know whether we *now,* while deliberating, believe that p.[15]

And we *do* have a theory of how, from within the first-person perspective, we deliberate about whether we believe some proposition. The central insight comes from the work of Gareth Evans, who observes that when we ascribe beliefs to ourselves, we survey, not facts about our psychology, but facts which bear on the truth of the content of the belief we are ascribing. In first-personal theoretical deliberation, the question of whether we now

---

[15]One may worry that this argument is relying a great deal on what is meant by 'now.' First, I think it is quite reasonable, as a matter of interpretation, to think that in many cases what we mean by 'now' is not indexed to the time of the asking, but to the answering. Suppose, e.g., I ask you what time it is; when you answer, you answer with the time of the answering, not the time of the asking. (for a fuller defense see Boyle (2011)). More crucially, however if we think of the 'now' in any other way, our belief that p will no longer really be self-fulfilling. Just like the belief I form about whether I win the match will not be self-fulfilling long after the match is done, though it may have been self-fulfilling at some earlier time. If the Millionaire said, I will give you a million dollars if you believed a minute ago that you would be a millionaire, the case is a much simpler, and less interesting, one.

believe that p is *transparent to* the question of whether p.[16] So in determining whether we believe that p, we must determine yet one further question: whether p.

How, then, are we to deliberate about whether we will be a millionaire? Since, for self-fulfilling beliefs, whether p is so depends on whether we belief that p, determining whether we will be a millionaire requires our opening up a new sub-question: 'whether I believe that I will be a millionaire.' But by transparency, to determine whether we believe that we will be a millionaire, we must direct our gaze outwards, and determine whether we will be a millionaire. This, in turn, requires us to determine whether we believe that we will be a millionaire, and so on ad infinitum.

Thus it now seems our deliberation ends, not in bootstrapping, but circularity. We appear not to arrive at a negative answer about whether p, but instead to fail to come to any answer at all. What, in such circumstances, is a believer to do?

Pretty clearly, it is impermissible for us to adopt the belief that we will be a millionaire. To do that would require that our evidence support a positive answer to the question 'whether p.' But, in following our evidence, we've failed to come to any answer at all.

Nor, contra the bootstrapping objection, is it permissible for us to believe that we will not be a millionaire. For to do *that* requires that our evidence support a negative answer to the question 'whether p.' But in following our evidence, we've failed to come to any answer at all.

One might think that this second step is too quick. Perhaps our failure to make up our minds can be leveraged to into an argument for disbelieving that we will be a millionaire:

---

[16] Evans (1982). Such transparency accounts are of course controvertial, though I shall not get into that controversy here. For a detailed account of the transparency phenomenon, see Moran (2001); Boyle (2011a); Byrne (2011)

the fact that we cannot make up our mind as to whether we will be a millionaire shows that we do not believe that we will be a millionaire (since failing to make up ones mind about p is a way of not believing that p). If we don't believe we will be a millionaire, we won't, so we should disbelieve. We would, in effect, have re-created the bootstrapping argument once again.

But this new objection faces the following dilemma. The train of reasoning, from the fact that I cannot make up my mind about p to the conclusion that not-p, may be part of the same deliberation in which I fail to make up my mind, or it may be part of some new deliberation. If it is part of a new deliberation, then my failure to make up my mind previously is no reason to think that I *now* fail to make up my mind. After all, that is exactly what the present deliberation will determine.[17] If, on the other hand, this second bit of reasoning is part of the same deliberation, my conclusion will be in conflict with my premises. For we have started by assuming we cannot make up our mind, and use this very fact to end up making up our mind that not-p.[18]

If neither belief nor disbelief are supported by the evidence, a natural third possibility is that, since we cannot decide whether p is so, the proper doxastic attitude to take toward a self-fulfilling belief is one of suspended judgment.

But if suspended judgment is to be a genuine propositional attitude, the claim that one should suspend judgment if one has insufficient evidence for or against a proposition must be more substantive than the claim that we should suspend judgment so long as we are in the position of being unable to establish p or not-p. Rocks, after all, can be in the position

---

[17] I am indebted here to Matthew Boyle (2011)'s argument: 11 – 12.

[18] A similar line of reasoning applies to the suggestion in Peels (2015):13 that we can leverage our awareness of the fact that our deliberation is inconclusive to conclude that we *will* be a millionaire.

of being unable to establish p or not-p, as can someone who has never thought about p.[19] As a genuine doxastic attitude, suspended judgment toward p should involve something like the taking of one's evidence to be insufficient to support either a proposition or its negation. And however this 'taking' is understood, it will at a minimum require, as with any other doxastic attitude, that we at least settle for ourselves the question of whether we will be a millionaire to the extent that we can have some stance on how our reasons ultimately play out. If the deliberation never concludes, we are not in the position of having determined that our evidence is neutral or counter-balanced toward the proposition, we are in the position of having failed to make up our minds about 'whether p' at all.

But if, like belief and disbelief, suspended judgment is also an inappropriate response to the circular deliberation, we have run out of candidates. So what *are* we to think about whether we will be a millionaire? I think that our investigation so far lends support to a fourth and final possibility, that self-fulfilling beliefs may be beyond the scope of our evidence altogether. In cases of self-fulfilling belief, where the evidence for the object of the deliberation depends on the outcome of the deliberation itself, our evidence under-determines what we are to believe.

§IV. Conclusion

In this chapter, I have examined the argument that the fact that a belief that p is self-fulfilling makes believing p impermissible, and claimed that this bootstrapping argument fails to hold up under stricter scrutiny. The objection rests for its plausibility on a particular function-picture of theoretical deliberation. I have argued that self-fulfilling beliefs can help us see that this picture is inadequate for a general understanding of theoretical deliberation. Because it is unable to take into account the fact that the outcomes of our deliberation can

---

[19] See Friedman (2013)

36

count as evidence for the proposition we are deliberation about, the function model is unable to account for cases where what we believe affects the truth of what is believed.

Finally, I have argued that with a better picture of theoretical deliberation, we can see that for self-fulfilling beliefs, our evidence does not dictate disbelief, but rather fails to dictate any appropriate doxastic attitude. The principle 'proportion your belief to the evidence" thus underdetermine what we are to believe with respect to self-fulfilling beliefs.

So what ought we to take from this exploration of self-fulfilling beliefs? One upshot involves the debate between the the pragmatist and the evidentialist. It turns out that pragmatism may not involve any direct disagreement with evidentialism after all. For insofar as the pragmatist limits herself to cases of self-fulfilling belief, it may be that the pragmatist norms are limited to cases that which fall outside the reach of evidentialist norms altogether. Looking closely at self-fulfilling beliefs reveals that the conventional picture of this debate may need to be re-evaluated.

But attending to the case of self-fulfilling beliefs also reveals a deeper, more general issue with our standard picture of epistemic rationality and the norms and nature of theoretical deliberation, which is otherwise easily overlooked.

Much of the discussion of epistemic rationality takes as its paradigm cases in which the belief in question is a belief in some independent truth, so that the role of our theoretical reasoning is to get us on to some independent fact of the matter, by following our evidence.

Such a point may seem like a platitude, and it may seem to follow from a straightforward realist picture of truth. For realism is often glossed as the position that there is an objective, mind-independent world out there to track. But we are not Cartesian egos, apart from the world and observing it. We ourselves are part of the world we are trying to represent, as our our beliefs. For that part of the world made up of or effected by our

37

beliefs, what we believe will affect what the objective world is like. Sometimes our beliefs don't simply track the truth, they create it.

If we fail to notice this point, and limit our attention to belief-states downstream from the parts of the world they purport to represent, it will be natural, as in the function model of theoretical deliberation, to think that the relevant considerations which bear on whether to believe will be independent facts about how the world actually is. And this will make it natural to think that the only appropriate epistemic norms will involve ones evidence, indicating the truth of the proposition.

Once we consider self-fulfilling belief, we can see that there are certain propositions where deliberation falls outside of the scope of our evidentialist norms. For when belief is self-fulfilling, since there is no independent fact, there will be no such evidence of that fact for us to find.

If we are to reason about those parts of the world, we will need a richer picture of epistemic rationality to accommodate it, and a more ecumenical approach to the the sorts of reasons and inferences which might appropriately fit into a picture of epistemic rationality anchored in the aim of truth.

Chapter 3: Epistemic Freedom Revisited

Consider again the following example: an eccentric millionaire comes to your door with the following credible offer: she will give you a million dollars if you believe that you will be a millionaire.[1] Should you believe that you will be a millionaire?

In the previous chapter I argue that an agent attempting to answer this question faces a distinctive puzzle. If, in the millionaire case, all a rational believer has to go on when deciding whether to believe that she will be a millionaire is the evidence she has available at the time she begins deliberating, she will find herself faced with an apparent deliberative gap. Since whether she will be a millionaire is a matter of whether she believes that she will, and since whether she believes that she will be a millionaire is a matter of how she proceeds in this very deliberation, there will be no appropriate doxastic attitude for the deliberating agent to take toward the proposition that she will be a millionaire, when that proposition is self-fulfilling. How ought such an agent proceed in her deliberation?

In this chapter, I want to examine a sustained attempt to answer this question, proposed by David Velleman. Velleman attempts to resolve the problem by advocating a form of "epistemic freedom" with respect to self-fulfilling beliefs: since whatever the agent believes will be true, the agent "may expect whatever… he wants."[2]

In section one, I will reconstruct a general account of Velleman's view. I will argue that two central points underpin Velleman's concept of epistemic freedom. First, that cases of self-fulfilling belief constitute cases of *epistemic permissivism* – cases in which an agent may permissibly adopt either of two different doxastic attitudes toward the same proposition.

---

[1] This example comes from Johnson (1998). For similar cases, see Velleman (1989) or Reisner (2013).

[2] Velleman (1989): 66

Second, that in cases of epistemic permissivism, it is possible for us to believe at will, i.e. to believe directly, on the basis of practical reasons.

In sections two and three, I separately examine and argue against each of these points. In section two, I argue against the possibility of epistemic permissivism. I argue that though it is frequently the case that one's practical reasons can permit each of two contrary actions, theoretical reason is structured such that one's epistemic reasons never permit us to adopt either of two contrary doxastic attitudes toward the same proposition. In section three, I argue that even if we accept that there are circumstances where theoretical reason permits two contrary doxastic attitudes toward a proposition, it will not follow that we will thereby be capable of believing at will.

I will thus argue, contra Velleman, that self-fulfilling beliefs fail to constitute an important exception to the widely accepted claim that believing at will is both impossible and epistemically impermissible. Why then revisit the notion of epistemic freedom? Because as the ensuing discussion will make clear, though it is widely accepted that believing at will is impossible, it is not at all obvious *why* self-fulfilling beliefs should not constitute an important counter-example; understanding how Velleman's argument goes astray requires a careful examination of the aim and structure of theoretical reason. This investigation of epistemic freedom will thus have broader philosophical implications: by attending closely to the special case of self-fulfilling beliefs, we can gain important insights into the nature of theoretical reason more generally.

In particular, I will argue, attending to the case of self-fulfilling belief reveals the importance of carefully distinguishing between the aim of acquiring a true belief and the aim of believing what is true. While one cannot usually fail to establish that one will acquire a true belief without establishing the truth of the believed proposition, in the case self-fulfilling

belief the two can come apart. I will argue that insofar as the aim of belief has to do with determining whether the believed proposition is true, it will be both impossible and impermissible to believe for pragmatic reasons.

§I. Self-Fulfilling Belief and Epistemic Freedom

On Velleman's view, what licenses you to form self-fulfilling beliefs, such as the belief that you will be a millionaire, is the complex fact that as soon as you are in the state of believing that you will be a millionaire, the fact that you believe that you will be a millionaire will provide you with good evidence to support your belief. In order for this fact about your future evidence to license your belief, we must understand the evidential requirements for belief broadly: it can be rationally permissible for an agent to adopt a belief, not only when she already has evidence for the truth of the belief, but so long as she *will* have evidence for the truth of the belief, once held.

Velleman defends this broader interpretation of the evidential requirements by showing how such an interpretation is consistent with what he sees as the underlying purpose of epistemic norms: that they help a believer attain the epistemic aim of true belief. Says Velleman:

> This purpose may well require a rule that one shouldn't retain a belief unless one has evidence of its truth. But does it require a rule that one shouldn't *form* a belief without *prior* evidence? I say no – at least, not if one has evidence that the belief would be true if one formed it. *Why would rules designed to help one arrive at the truth forbid one to form a belief that would be true?*[3]

Velleman here makes what I think to be an important observation – while following your evidence is usually a sure method for arriving at the truth, it is not the only sure method. In cases where the outcome of our deliberation helps determine the truth of our belief, that belief may be formed absent prior evidence of its truth, but not thereby be formed

---

[3] ibid, 63. (Emphasis added)

irrespective of its truth. Velleman is right that rules "designed to help one arrive at the truth" can allow for more latitude than stricter formulations of evidentialism suppose.

But while Velleman may have shown that beliefs arrived at ahead of the evidence may be nonetheless reliably connected to the truth, he has not yet provided an explanation of how we, as believers, when faced with the millionaire's challenge, are supposed to determine what to believe, if not by following the evidence. For while we may want to know, of some already existing belief, how and whether that belief is reliably connected to the truth, this is not the only role for an account of theoretical rationality to play. We might also want to know how we, as rational believers, ought to arrive at those beliefs in the first place. We might be concerned, from within the perspective of the agent engaged in theoretical deliberation, with the sorts of reasons she might rationally employ in her reasoning and the sorts of norms which might rationally guide her in that reasoning.[4] And as an account of how we, as rational believers, ought to deliberate, Velleman's theory faces two obstacles, which I shall mention briefly here and then proceed to discuss in more detail.

In providing such an account, one's theory of rationality should be the kind of thing that a rational agent can follow. It must involve the sorts of norms that can guide or advise the agent in her deliberation.[5] Just as good practical guidance must be guidance an agent can act upon, good theoretical guidance must be guidance an agent can believe upon. But the advice: "believe what will be supported by your future evidence" does not obviously appear to be the kind of advice an agent can follow. For one thing, at the time one is initially engaged in deliberation about what to believe, one does not yet have that evidence available to employ as a reason to believe. The advice requires a believer *run ahead* of her possible

---

[4] Though there may be other important epistemological projects, this seems at least *one* important epistemic project with which we might be concerned. See Goldman (1980) or Berker (2013) for further discussion.
[5] See, e.g. Arpaly (2003). For discussion of belief, in particular, see Wedgewood (2002).

reasons for believing. For another thing, in at least many cases when a belief is self-fulfilling, both a belief that p *and* a belief that not-p will be such as to be supported by evidence, were one to believe them. The advice thus appears to *underdetermine* what an agent is to believe.

a. The Attraction of True Belief

If Vellman's theory is to be an account of how, from the perspective of the agent, one ought to engage in theoretical deliberation, it must provide reasons which the agent might be able to employ, in her theoretical deliberation, to come to her belief.[6] And these reasons cannot be the evidence that would support the belief only once it is believed, because, as the agent does not yet have the belief, she cannot yet employ that evidence in her initial deliberation.

Here is how Velleman first addresses the question of how the agent is to arrive at a belief, despite lacking the evidence for the belief at the time the belief is formed:

> forming a self-fulfilling belief isn't really "running ahead" of the evidence: it's running toward the evidence – that is, toward evidence that will consist in the belief itself… what would attract him to a conclusion for which the evidence still remained to be completed? What attracts him, in most cases, is simply the prospect of making the conclusion true by jumping to it.[7]

The idea, I take it, is that your reason for believing that you will be a millionaire just is the fact that you will have evidence for holding the belief once you believe it. And this reason, unlike evidence you lack until the belief is held, can be available to guide you in acquiring the belief. For while you may not be able to adopt a belief on the basis of evidence you lack until

---

[6] This is not to say that every time an agent permissibly adopts a belief, she must be engaged in high-level reasoning of this kind.

[7] Velleman (1989): 64

the belief is acquired, you might still be able to treat the fact that you will have evidence later (a fact which you have access to now) as a reason to adopt the belief.[8]

But though the fact that you will have evidence for holding the belief once you believe it can be a reason for which you believe that you will be a millionaire, it is still not clear how this reason can suffice, on its own, to explain your believing that you will be a millionaire. This is because you have the very same reason to judge that you will not be a millionaire. For if you believe you will not be a millionaire, you won't. And once you believe you won't, the fact that the belief is self-fulfilling will give you evidence to support your belief that you won't be a millionaire. So if all that attracts you to an outcome is the prospect of making the conclusion true, we still lack an explanation of why you jump to the conclusion that you will be a millionaire, rather than jumping to the conclusion that you will not be a millionaire.

The fact that you will have evidence to support a belief, though it can be the kind of reason employed by a deliberating agent, can be at most a beginning of the story of why you believe that you will be a millionaire, not the end of it. And this is troubling, since no story in terms of further epistemic reasons – either having to do with the evidence you now have, or the evidence you will have – appears forthcoming.

b. Practical Reasons and Self-Fulfilling Beliefs

It is here that it first becomes important that the example we are using – that of being a millionaire – is an example which involves a question about whether or not some *positive,* or desirable, state of affairs is so. Velleman agrees that more needs to be said to finish the story of how, in the face of symmetrical reasons, the agent believes that p, rather than

---

[8] One might worry that such an account will still involve an illicit bootstrapping, if your reason is the fact that you *will* have evidence for p once you believe that p, rather than the fact that you *would* have evidence for p were you to believe that p. This issue will resurface in section III.

not-p. His response is that what breaks the tie for the agent when she is deciding what outcome to believe is that she would prefer to believe that that she will be a millionaire – that is, that the agent decides between which of the two conclusions to believe on the basis of her practical considerations for preferring one or the other.

This may look like a surprising continuation, if what you were expecting was a continuation of the story of why you will, rather than won't, be a millionaire. Unlike the fact that your belief that you will be a millionaire will be supported by evidence, the fact that it would be *good* if you were a millionaire does not appear to bear any rational connection to the truth of your belief. So it is very hard to see how this can provide you with any further epistemic reason to believe that you will be a millionaire.

This, says Velleman, is because these practical considerations are *not* further epistemic reasons to believe that you will be a millionaire. But this is no problem, for while more needs to be said to explain how a rational agent jumps to one conclusion rather than the other, no more needs to be said about the epistemic *reasons* for which the rational agent comes to her conclusion. No more needs to be said about the reasons for which an agent arrives at the belief that p when the belief is self-fulfilling, because while the fact that a self-fulfilling belief that p will be true when believed does not give an agent reason to believe that p rather than believe that not-p, it can still provide the agent with reasons sufficient to show that a belief that p will be true. Unlike a sufficient *explanation* of how one comes to a belief or action, which must rule out the alternatives, a sufficient *justification* of why one believes or acts need not require one rule out all the alternatives.

Those who are unhappy with this answer, Velleman suggests, are reluctant because they have forgotten an important lesson from Harman, that:

> the pattern of reasoning is not its logic. The order of theoretical reasoning, for instance, is
> the order in which beliefs are formed, which is a psychological matter. But beliefs aren't

necessarily formed in order of logical dependence – premises first, conclusions following. Often we adopt a conclusion and then look for evidence, or reject evidence once we see which conclusion it supports. Logical order is something that we try to attain in our beliefs, but we don't attain it by forming beliefs in their logical order.[9]

If our completed reasoning shows that our belief will be reliably connected to the truth, no more needs to be said in terms of a rationalizing explanation. All that needs to be provided is a psychological story of how we jump to one conclusion over the other. The story is that we simply choose whichever conclusion we prefer to jump to, by bringing to bear in our reasoning practical considerations which show jumping that way to be choice-worthy. We decide at will what conclusion to believe, and then engage in theoretical reasoning to support it.

Velleman admits that this move is unconventional. That we prefer a certain outcome is not a further reason to think the outcome will happen. But, he thinks, we do not *need* any more reason – we already had sufficient epistemic reason to permit belief in either outcome. To demand, when either belief is epistemically permissible, that the explanation of how we arrive at one belief rather than the other be in terms of further epistemic reasons is to demand answers where no further answers exist. So far as theoretical rationality is concerned, as long as one has reasons which show a belief is reliably connected to the truth, no more needs to be said. And as the requirements of theoretical rationality are satisfied either way, nothing is stopping us from rationally believing at will, on the basis of practical reasons.

c. Doxastic Involuntarism

All that is left, says Velleman, is to show that this pattern of reasoning is psychologically possible. For having shown that we may believe at will, it remains to be

---

[9] ibid, 108; See also Harman (1973)

shown that we *can* believe at will in these cases. Velleman does this by responding to what he takes to be the main objection, from Williams, against the possibility that we can believe at will.

The argument, following Williams, is that it is at least partly constitutive of beliefs that they purport to represent their content *as true*. But if we, in full consciousness, acquire our belief through any means not directed toward the truth of the content, we would know we have no reason to think of the content as true, and so have failed to actually form an attitude which amounts to a *belief* about the proposition.[10] But as Velleman points out, this will not be a problem for self-fulfilling beliefs:

> because self-fulfilling beliefs cause their own truth, instead of being caused by it, their connection to the truth isn't threatened by the fact that they are caused by the subject's wishes. They remain reliably connected to the truth even if they are determined by what the subject wants to believe. This means that [the subject] can therefore think of himself both as believing what he likes and as believing only what's true, because what's true depends on what he believes.[11]

Velleman makes use of a lacuna in Williams's argument, first noted by Barbara Winters. Winters notes that what is incompatible with taking our representation as true is not that willed-beliefs are *caused* or *generated* by some mechanism other than truth-relevant considerations, but that, if we are conscious of the fact that we believe at will, these beliefs will not be *sustained* by truth-relevant considerations.[12] But since a belief need not be sustained by the same considerations by which it was generated, there is at least some conceptual space for attitudes which are brought about by some act of will, but which are sustained, not by will, but by truth-relevant considerations, and so still properly counted as beliefs.

---

[10] Wlliams (1973): 142

[11] Velleman (1989): 128-129

[12] Winters (1979): 246; See also Bennet (1990)

And, as Velleman notes, self-fulfilling beliefs are precisely the sort of beliefs which fall under this exception carved out by Winters. Though there may not be sufficient evidence with which to adopt the self-fulfilling belief, once the belief is adopted there will immediately be sufficient evidence to sustain the belief. And so someone might form such a belief at will, and then immediately sustain the belief on the basis of the evidence now available to her, and so have no trouble with taking the attitude she has adopted as purporting to represent the truth. And so Williams' argument, as formulated, gives us no reason to suspect that believing at will, at least for self-fulfilling beliefs, is not a psychological possibility.

d. Summary

In summary, Velleman advocates a form of *epistemic freedom* with respect to self-fulfilling beliefs: since the outcome will be whatever the agent expects, the agent "may expect whatever… he wants."[13] When faced with a self-fulfilling belief, you ought to adopt a belief on the basis of considerations which show that were you to believe that p you would have a true belief, and on the basis of considerations in virtue of which you might wish or want to believe that p.[14] For Velleman, this reasoning will involve an amalgom of practical and theoretical reasons. The first set of considerations suffice to show your belief that p will be true if you believe it, and so make your belief theoretically permissible, and the second set of considerations suffice to show what you believe choice-worthy, and so makes your belief practically rational.[15]

---

[13] Velleman (1989): 66

[14] The most plausible examples of this pattern of reasoning involve the (purportedly) self-fulfilling beliefs Velleman is most interested in: our beliefs about our own actions.

[15] In some cases, the same consideration or set of considerations might play the role of both your practical and your theoretical reason. Thus I might believe that I will be a millionaire, because if I believe that I will be a millionaire, I will be, where this consideration shows the belief both true and desirable.

I have argued that two points underpin this picture of epistemic freedom. The first point involves a form of epistemic permissivism we enjoy with respect to self-fulfilling beliefs. Our theoretical reasons – those considerations which show that a belief will be supported by evidence – can provide us with *sufficient* reason to permit the self-fulfilling belief that p, even though they do not provide us with determining reasons to believe that p, rather than believe that not-p. The second point is that this phenomenon of epistemic permissivism makes possible a limited form of doxastic voluntarism. When whatever we believe with respect to some proposition is epistemically permitted, we are free to believe at will on the basis of further practical considerations.

I think there are problems with both points. First, I will argue that we have strong grounds for thinking that in theoretical deliberation, unlike practical deliberation, sufficient reasons must also be determining reasons – reasons to believe that p *rather than* not-p. In consequence, there can be no cases of epistemic permissivism. Second, I will argue that even if there are circumstances where our reasons can permit two contrary doxastic attitudes toward a proposition, it would not follow that we could believe at will whatever we like. Epistemic permissivism would not make possible the limited kind of doxastic voluntarism which Velleman's picture of epistemic freedom requires.

§II. Self-Fulfilling Belief and Epistemic Permissivism

To recapitulate, Velleman thinks that your reason for believing that you will be a millionaire is that the belief will be supported by evidence, once you believe it. And he thinks that this is *sufficient* reason for which you can come to believe that you will be a millionaire because, by showing your belief to be reliably connected to the truth, it shows *now* that what you believe *will be* true. If this is a sufficient reason to believe that you will be a millionaire, then you also have sufficient reason to believe that you will not be a millionaire, since the

fact that a belief is self-fulfilling also shows that this contrary belief will be reliably connected to the truth as well. Velleman is endorsing a form of epistemic permissivism – your epistemic reasons permit either of two contrary doxastic attitudes toward the same proposition.[16]

Velleman's argument rests on the premise that whether some set of reasons is sufficient to permit a belief is a matter of whether they suffice to show that the agent has met her epistemic aim of true belief.[17] Roughly: an agents reasons for believing that p are sufficient just in case those reasons are sufficient to show the belief that p true. But so stated, this formulation contains an important ambiguity. The "belief" whose truth must be shown might refer to a belief-state, or it might refer to *what* is to be believed – a proposition. Accordingly, in adopting a belief that p, there are two different possible sets of considerations that might be required for an agent to satisfy the aim of belief: considerations sufficient to show that the agent will acquire a true belief and considerations sufficient to show that the believed proposition is true.

It is rare that this ambiguity is important. A set of considerations cannot, in normal circumstances, be sufficient to show that you *would be* believing something true without also being sufficient to shown that the proposition to be believed *is* true.[18] In fact, the best way to

---

[16] In this, Velleman has been joined by Peels (2014); Reisner (2013) and Foley (1999). Wedgewood (2013) endorses a related claim for the *correctness* of belief. As we will see, however, it is not obvious, from the fact that a belief that p and a belief that not-p would both be correct, that a belief that p and a belief that not-p would both be permissible, in the sense at issue.

[17] Velleman (1989): 63. For further discussion of the Aim of Belief, see Williams (1973), Velleman (2000), Wedgewood (2002), Hieronymi (2005), among many others.

[18] I contrast what *would be* true with what *is* true, because I assume that the truth of a proposition is eternal (though, in the case of self-fulfilling belief that p, p is eternally true only because you will eventually believe it). Nothing hangs on this assumption, however. If one accepts some thesis about the indeterminacy of future contingents, one can instead speak of the contrast between what *would be* true and what *will be* true.

determine whether you will have a true belief-state is usually by way of determining whether the believed proposition is so.

This is because, usually, the truth of a proposition is independent of an agent's belief in that proposition. This means that if p is false, it follows that you would also, in believing that p, believe something false. But when the truth of a proposition depends on whether you believe the proposition, this no longer follows. For were you to believe that p, the truth value of p might thereby be different than it is in the actual world, where you will not.

This can be seen most clearly by looking at what I will call 'self-defeating beliefs.'[19] Consider a possible Moorian proposition of the form: q but I don't believe that q. Such a proposition may be true in the actual world, though an agent may not arrive at a true belief were she to believe it, because in such counter-factual circumstances, she would also believe q, and so the Moorian proposition would be false.

Conversely, for some self-fulfilling belief that p, it may be both that the proposition p is false (because, actually, you do not and will not believe it) and also that you would still, in believing p, believe something true, because in the counter-factual circumstances where you believe that p, p would have been true, in virtue of your having believed it.

In such cases, that p is so no longer guarantees that, in believing that p, you would be believing a truth, or that, in believing not-p, you would be believing a falsehood. This has two important consequences. First, it means that, in such circumstances, it may be that you no longer need to determine whether p is so to determine whether, in believing that p, you would be believing something true. For you may be in a position to determine that your belief would true regardless of whether the proposition you are believing is true.

---

[19] These are what Gombay (1988) refers to as the "paradoxes of counter-privacy" or, more recently, cases Bykvist and Hattiangadi (2013) refer to as deliberative 'blindspots.'

Second, it means that, given the appropriate counterfactuals, you may find yourself in the position of having both reasons which show that, in adopting the belief that p, you would adopt a true belief, and which show that, in adopting the belief that not-p, you would adopt a true belief. Because while two such beliefs cannot not both actually be truly held, it could nonetheless be that either belief, were it held, would be held truly. In these cases, by establishing that whatever you might come to believe, you *would* believe something true, you can establish that whatever you do come to believe, you *will* believe something true.

And when a belief that p is self-fulfilling, you are in precisely such a position. The fact that a belief that p is self-fulfilling is sufficient to show that in believing p, you would be believing something true, and it is also sufficient to show that in believing not-p, you would be believing something true.[20]

Consider now the prospects for epistemic permissivism on the second reading of the aim of belief, on which sufficient reasons for belief must suffice to show that the proposition to be believed is true. Could the fact that a belief that p is self-fulfilling be sufficient both to show that p is true and to show that not-p is true? I think the answer is surely no. Nothing is sufficient to show that, since p and not-p cannot both be so. Since p and not-p cannot both be so, your reasons, insofar as they are sufficient to show p true, will thereby be sufficient to show not-p false.

This is importantly different than the case of practical deliberation about what to do. In practical deliberation, an agent may find herself in circumstances in which two contrary

---

[20] Moreover, you can establish that your belief will not only be true, it will also satisfy the sorts of modal anti-luck conditions often associated with knowledge. Your belief will enjoy both *safety* and *sensitive*: in all nearby possible worlds where you hold the belief, the belief will be true, and in all nearby worlds where the belief is false, you fail to believe it. In the usual examples of safety and sensitivity of course, the counter-factual obtains because you are the sort of believer who discriminately tracks the truth. Attending to the case of self-fulfilling belief reveals that self-fulfilling beliefs can also enjoy these modal properties because the truth is sensitive to your believing it.

actions involve equally good outcomes. Drinking a glass of lemonade or a glass of iced tea might both be equally good ways to quench your thirst, and so the fact that you are thirsty might still give sufficient reason to permit drinking the lemonade or the iced tea, even if all the relevant facts were available. The fact that sufficient reasons for belief, unlike sufficient reasons for action, must also be determining reason can be explained in part because of a fundamental difference between the good and the true. While two incompatible states of affairs, p and not-p, might both be equally good, they cannot both be equally true. The sorts of cases of permissivism one finds on the practical side are precluded on the theoretical side by the law of non-contradiction.

And so, if we accept the second reading of the aim of belief, sufficient reasons for the belief that p must also be determining reasons for believing that p, rather than not-p. For in order for a set of reasons to show that the world is a particular way, they must also thereby be sufficient to rule out it not being that way.[21] In consequence, the fact that you will answer correctly however you answer the question whether p, because it is not yet *sufficient* to distinguish which of p or not-p is so, cannot yet be sufficient to show either p or not-p true. More simply: knowing that whatever you answer will be correct is not the same thing as having an answer.

When what we believe affects what will be so, the question *whether, in believing p, I will believe something true* is not transparent to the question *whether p*. The fact that a belief is self-fulfilling can give reasons which are sufficient to answer the first question, by showing that you as a believer are reliably connected to the truth, without being sufficient to answer the

---

[21] This is, at the very least, a condition on what is required to rationally settle the question of whether p. I suspect, though I leave open, that this condition is even stronger. It seems to me that it is not only impermissible, but impossible, to settle affirmatively the question whether p without thereby also settling, negatively, the question whether not-p: theoretical questions are fundamentally of the form "whether or not p"

question of what is so. For though your beliefs are reliably connected to the truth, it is not because they are tracking some independent truth in the world, but rather because they are determining how the world will be. So in knowing a belief will be reliably connected to the truth, you do not yet know how the world is.

What Velleman's argument from the aim of belief reveals, then, is not that self-fulfilling beliefs are permissible, but rather the remarkable fact that sometimes the aim of acquiring a true belief and the aim of believing what is true can come apart. Insofar as the aim of belief is concerned with our accuracy as believers, the fact that a belief is self-fulfilling can permit both the belief that p and the belief that not-p, for it shows that in believing either, we will believe truly. But insofar as the aim of belief is concerned more directly with our theoretical reasoning about the believed state of affairs, reasons which are sufficient to believe that p must give us reasons for answering the question 'is it the case that p?' one way rather than the other. And so the fact that a belief is self-fulfilling, rather than permitting both the belief that p and the belief that not-p, will on its own be insufficient reason to permit either one.

Velleman's permissivism thus rests crucially on the fact that we take the first reading of the aim of belief. If instead the second reading of the aim of belief is the right one, on which sufficient reasons to believe that p must be sufficient to show p true, then there will be no epistemic permissivism. For sufficient reasons for believing some proposition, p, must also be *determining reasons* – reasons to believe that p rather than not-p.

And while I do not have the space here to argue fully for the second reading, I suspect that once we are able to see these two interpretations come apart, much of the plausibility of Velleman's position evaporates. For once we see that there can be a difference between determining whether an agent will form a true belief, and determining whether the

proposition the agent is forming a belief about is true, we can see that what is central to belief is the latter.

For believing is not a special sort of intentional action, distinctive only in in that it must involve the goal of believing what is true. If it were, we would have no means of distinguishing belief from silently asserting, or from other mental activities like guessing.[22] Rather, belief is the kind of state constituted by or resulting from processes, like theoretical deliberation, which function to track the truth. Such processes are not concerned with whether we are being accurate believers, but with discovering *how the world is* (or at least the relevant bit). And insofar as this is our purpose as believers, we cannot, as rational agents engaged in deliberation, be satisfied with knowing that our beliefs will be true, whatever those beliefs turn out to be. We need to determine the truth of the proposition about which we are believing.

§III. Self-Fulfilling Belief and Doxastic Voluntarism

In the previous section, I argue against the claim that when a belief is self-fulfilling, we may permissibly believe either p or not-p. But let us grant for the moment that the fact that a belief that p is self-fulfilling *can* provide an agent with epistemic reasons which are sufficient to permit both the attitudes of belief and disbelief toward p. Velleman will then have shown that we enjoy a certain degree of rational latitude with respect to self-fulfilling belief.[23] But he has not yet shown that we can thereby believe whatever proposition we want.

---

[22] A contrasting case put to use to great effect by David Owens in Owens (2000) and Owens (2002). Guessing is a type of intentional activity, type-individuated in part by the fact that one's intention in guessing is to guess truly. Part of what it is to be guessing is to have adopted a standard of correctness involving guessing a proposition only if, in doing so, your guess will be true.

[23] This is the situation Velleman has in mind in Velleman (2000b) which he glosses in the following way: where there is "no particular way we must describe the future as turning out, in order to describe it correctly… epistemic freedom amounts to the fact that he would be correct in predicting events that aren't going to occur.

For when your epistemic reasons permit two attitudes, it's not straightforward that you can thereby choose one for any reason you please. In fact, it's not entirely obvious *what* a rational agent is supposed to do when her reasons are sufficient for two contrary attitudes, even in the practical realm where such cases are less exotic.

In this section, I will take up the question of how an agent ought to engage in theoretical reasoning when her reasons are so counter-balanced. Even though (as I have argued) there may be no such cases, thinking about what a rational agent would have to do were her epistemic reasons so counter-balanced can help us gain new insight into the nature of theoretical reasoning, just as thinking about similarly situated practical agents can help us sharpen our theory of practical reasoning.

I will pursue this question by looking at analogous cases on the practical side. Looking at these practical cases reveals important limitations of how cases of epistemic permissivism might ever be resolved. In the remainder of this paper, I will argue that however these cases are to be resolved, it will not be in a way which makes possible anything like the familiar notion of believing at will, where we decide what to believe directly on the basis of practical considerations.[24]

a. Moral Permissivism

One place to begin is by looking at cases of *moral permissivism*. These are cases where two or more of some conflicting set of potential actions are both morally permitted, but no one action is morally required. In such cases, most people do think that an agent might then choose one action over the other on the basis of other non-moral reasons. In particular,

They aren't going to occur only because he isn't going to predict them; and so if he did predict them, they would occur, and his prediction would be true." (36-37)

[24] In the proceeding, I will rely on a relatively general conception of believing at will, without delving too deeply into the details of what believing at will must involve. For an overview of the ongoing debate about these details, see Bennet (1990); Setiya (2008); and Hieronymi (2009).

most people think that an agent can choose one action over the other on the basis of which action she more prefers. You can rationally decide to volunteer at the Red Cross rather than the soup kitchen, for instance, because of your preference for learning first-aid over cooking.

In cases where our moral reasons are counter-balanced, such prudential considerations can fill in the gap and allow us to decide on one action over another. And so some have thought, by analogy, that if both the belief that p and the belief that not-p are epistemically permitted, we can believe one over another on the basis of non-epistemic reasons, because we prefer that belief more than the other.[25]

Still, when moral reasons permit two or more actions, an agent cannot just decide to do one action over the other on the basis of *any* reason they please. You cannot, when your moral reasons are tied, decide to volunteer at the hospital because you love Sophocles.[26] Your reason for acting, though non-moral, must still help make your action *intelligible*.[27] In the case of practical deliberation, a consideration makes an action intelligible by showing the action to be favorable, or choiceworthy, in some way. And the reason you can make recourse to prudential reasons when the moral reasons are counter-balanced is that prudential considerations, like moral considerations, are the kinds of considerations which bear (or can be taken to bear) on your question of what to do, by showing some possible action to be choice-worthy.

---

[25] See, for example, Reisner (2009)

[26] Raz (2001): 215

[27] This point does not require any strong claim about our acting "under the guise of the good." The point should hold so long as this notion of intelligibility is strong enough that it require a practical reason must bear, for the agent, *some* minimal rational connection to the action, and by parallel, that theoretical reasons must bear some connection to the object belief. This should be acceptable, even to skeptics of 'guise of the good' theses, like Velleman (2000c) or Setiya (2007).

Similarly, in the case of theoretical deliberation, if your evidence is counter-balanced, you can't decide to believe one proposition over the other on the basis of any reason you please. You would need some other category of epistemic reasons – some other category of considerations which still make your believing intelligible by bearing (or being taken to bear) on the truth of that belief. And considerations which show one belief to be preferable to another are not such considerations. And so, while such practical considerations can be your reasons for acting when your moral reasons are counter-balanced, they cannot be your reasons for believing when your evidence is counter-balanced.

b. Buridan Cases

The better practical analogue for Velleman is a case where two actions are permitted because *all* of an agent's practical reasons are counter-balanced, not just her moral reasons. How, in practical deliberation, do we decide what to do in such cases? On one prominent tradition, what breaks the tie is a special act of will. When our reasons are perfectly counter-balanced, we experience the *liberty of indifference*, and may act however we please.[28] I suspect Velleman thinks that in the case of an epistemic analogue we may believe however we please. Just as we break the tie in the practical case via an exercise of the will to act, we should break the tie in the epistemic case via an exercise of the will to believe.

But once again, the analogy breaks down when we look more closely at the examples on the practical side. On the practical side, such cases are commonly referred to as "Buridan Cases," after the example of Buridan's Ass, torn between two equally appetizing bales of hay. Such Buridan cases divide into two sorts, those involving imperfect and perfect information. In the first imperfect class of cases, a practical agent is indifferent between two options because, given her present information, those two options appear equally promising.

---

28 For a survey of historical thought on the issue, see Rescher (1960)

Suppose you come to a fork in the road, know that one path leads to your goal, but do not know which. Since you certainly won't get to your goal if you do nothing, you have reason to pick each of the two paths. But, though you know taking one fork will involve a better outcome, since you don't know which fork it is you have no available reason to prefer one over the other.

By contrast, the second class of "true" Buridan cases involve situations in which an agent has all the relevant information, and still has equal reason to choose both options. Suppose you learn both paths lead to your goal. Here you must still choose one path over the other without having more reason to choose one path over the other, not because you are missing any relevant information, but because both choices involve what are, in fact, equally good options.

I suspect that part of the reason epistemologists have traditionally been unconcerned with Buridan cases has to do with this division.[29] In cases of imperfect information, epistemologists have held that one is never faced with two permissible attitudes to choose between, because the attitude of withholding is always available as an option in theoretical deliberations in a way that it is not available in at least some practical deliberations.[30] You may find yourself forced in a situation in which you must *take* one path over the other, but you do not have to *believe* that your path is any more likely to end at your goal. In cases of imperfect information, when one's epistemic reasons are counter-balanced, the obligatory doxastic attitude is that of suspended judgment.

---

[29] A notable exception is Reisner (2013), who deals with the possibility of epistemic Buridan cases explicitly, in just these sorts of situations.

[30] A point made in Feldman (2000): 679-682 and Feldman (2006), contra James (1895).

And epistemologists have assumed that there was no theoretical analogue to the "true" Buridan cases because when one has all the information, ones reasons will never show two contrary belief states equally reasonable. For unlike practical cases, where two mutually exclusive outcomes can both be equally good, two mutually exclusive states of affairs cannot both be true.

But as we have seen in the previous section, in the special case of self-fulfilling belief one *can* find oneself in deliberative situations where either of two mutually exclusive doxastic states will be true if you to adopt them. Consider, by way of practical analogy, a special case where God will create the hay in whichever direction the donkey chooses.[31] While the outcomes in turning left or right in such circumstances would be radically different (one direction will have hay, the other will not), the outcomes of *deciding* to turn left or right would be the same. Self-fulfilling beliefs appear to be theoretical analogues to this special sort of "true" Buridan case on the practical side. Given these similarities, we might look to the practical responses to Buridan cases, and see what relevance they have to the case of self-fulfilling belief.

There are several theories of what a rational agent is to do in practical Buridan cases, but none of them, I will argue, are going to help us reach Velleman's position that when either of two beliefs are epistemically permissible, we can believe at will whichever we want. Moreover, attending to the practical case and its epistemic analogue will help show that a capacity to believe at will could never be the sort of ability which could help resolve a case of epistemic permissivism.

c. Practical Responses

---

[31] I am grateful to Calvin Normore for this helpful example.

I turn now to consider the possible practical responses to Buridan cases. One possibility on the practical side is that when our practical reasons are counter-balanced, we fail, or should fail, to choose anything. And while ethicists have tended to doubt that rational agents would be so asinine as to starve amidst an over-abundance of good choices, it might be less strange to think that a rational agent would be unable to make up their mind about what to believe when faced with an over-abundance of possible true beliefs. This option is of no use to Velleman however, since what Velleman wants is some story about how we *can* adopt a belief in the face of counter-balanced reasons.

Among ethicists who think that Buridan's Ass does not starve, there are broadly two lines of thought as to why. One generally Aristotelian line is that I chose to act under an action description for which my reasons are *not* counter-balanced, and some causal or explanatory mechanism decides the particulars.[32] Thus, I might decide to take *some* path, and while I will take either path A or path B, I do not choose to do so on the basis of some further reason, but rather do so on the basis of some purely causal process.

Of course, even if we accept the controversial thesis that we can perform the same action under a variety of descriptions,[33] this option is famously unavailable to us in the doxastic case. Belief contents are opaque, and if we believe that a state of affairs obtains under two different descriptions, we have two different beliefs. I cannot believe, for example, that there is treasure at the end of path A by believing there is treasure at the end of some path, in the way I can perhaps intend to take path A by intending to take some path.

---

[32] A position championed by Avverroes (1954), and more recently, by Donald Davidson (1963): 687-688.

[33] See on the one side, Anscombe (1957), on the other Davidson (1963).

The second line of thought is that when we are faced with counter-balanced reasons to act, we exercise a special power of the will to just decide upon one action or the other.[34] When I come to the fork in the road, rather than acting as the result of weighing the balance of my reasons, I just pick whichever action I like.

This may look more promising for Velleman as a practical analogue, since it appears to get him exactly what he needs. Just as you exercise your will to pick what to do when your reasons for acting are counter-balanced, you exercise your will to pick what to believe when your reasons for believing are counter-balanced. I suspect that this, or something close to it, is what Velleman has in mind in claiming we can believe at will. However, a closer examination of what must be involved in this special exercise of the will reveals such an analogy to be unworkable.

Typically when we speak of willing, we are speaking of willing *for reasons*. I act at will, for instance, when I choose to spend the afternoon at the park rather than the cinema because of my preference for the outdoors. When Velleman speaks of believing at will he must presumably have this sort of willing in mind if, on his picture, the explanation of how we come to one belief rather than another when our epistemic reasons are counter-balanced is that we decide what to believe on the basis of our preferences or practical reasons.

But the exercise of will posited in order to resolve practical Buridan cases does not involve, nor could it ever involve, this typical sort of willing. For in practical Buridan cases, the reason a rational agent needs some further mechanism is that all of her practical reasons have already come in to play, and have been insufficient to resolve the agent's deliberation about what to do. The whole point of positing an act of will is to allow the rational agent to

[34] A position championed by Al-Ghizali (2000), and more recently among Rescher (1960), Morganbesser and Ullmann-Margalit (1977), and Normore (1998). Reisner (2013) considers, and rejects, a similar account of theoretical reasoning for self-fulfilling belief.

resolve a deliberation by answering a question of what to do *without* recourse to any further reasons. Such a willing is sometimes referred to as a "picking" to distinguish it from "choosing" which is always done for reasons.[35]

So when we say that we are free to act as we want when our practical reasons are counter-balanced, it is in a very specific sense: it is not that we choose on the basis of our reasons, preferences, or desires – that's what we regularly do – but rather that we are willing to *make* that action the one that we prefer. Thus even if we could will to believe what we want in this special sense when our epistemic reasons permit belief in either of two contrary propositions, it would not amount to believing at will, at least in the familiar sense where we decide what to believe directly on the basis of practical considerations. We could not exercise this power to believe that we will be a millionaire for the reason that we would like to be a millionaire, because this would be an act of choosing what to believe for reasons.

So if cases of epistemic permissivism would allow for the possibility of believing at will, it would not be the kind of familiar believing at will, on the basis of practical reasons. But by thinking about what a theoretical analogue of picking would have to be like, we can also see that an act of will could never be the right kind of capacity to help an agent resolve an epistemic Buridan case. No power of the will, with our beliefs as object, could help resolve the Buridan-problem in which cases of epistemic permissivism place an epistemic agent.

We can see this by thinking about what a resolution to a Buridan problem must involve. Consider first the situation of the practical agent. A rational agent, engaged in practical reasoning about what to do, should be able arrive at a decision, and so act or intend. Often such an agent can accomplish this by weighing her respective reasons for

---

[35] A nomenclature originating with Morganbesser and Ullmann-Margalit (1977).

acting. In practical Buridan cases, the agent faces a problem. Since her reasons are sufficient for two different actions, she requires some further capacity to resolve her otherwise intractable deliberation about what to do, and come to a decision about how to act. A power of the will to pick how to act, without picking for further reasons, would be the right kind of power to solve the problem. So willing would help a rational agent conclude her practical deliberation, allowing her to come to a decision about how to act.

Epistemic Buridan cases, in contrast, involve a rational epistemic agent, engaged in theoretical deliberation, who needs to resolve her reasoning about *what is so* and thereby come to some doxastic attitude. Often, an agent can accomplish this by weighing her respective reasons for believing. In epistemic Buridan cases, since her reasons are sufficient for two different doxastic attitudes, she would require some further capacity to resolve her theoretical deliberation and so come to a belief about what is so.

It may seem as though a power of the will to pick what to believe, without believing for further reasons, could be the right kind of power to resolve such an intractable theoretical deliberation, just as a power of the will to pick what to do can resolve an intractable practical deliberation. Thinking carefully about what is needed to resolve a theoretical deliberation, however, reveals such a capacity to be inadequate.

We can see why this is so by again distinguishing carefully between acquiring a true belief in some proposition and concluding some proposition true. Theoretical reasoning, insofar as it is reasoning about *what is so*, requires, for its resolution, the latter. And so if we need to posit an extra power to resolve an otherwise intractable theoretical deliberation, it will need to be a power which allows us to conclude that p is so.

But a power to will yourself to have one belief or another – even a power to will yourself to have one of what you know would be two true beliefs – is simply the wrong kind

of power to effect such a resolution. Exercising such a power does not resolve your theoretical deliberation, it circumvents it. To resolve your theoretical deliberation, you would need a power, not to will one belief or another, but rather to take a proposition or its negation as true. For it is only by taking a proposition or its negation as true that an agent will have done something amounting to determining whether p, resolving her reasoning.[36] Just as it is only by taking an action as to be done that we would be able to resolve a practical deliberation about how to act.[37]

And so an act of will could never help an agent in such circumstances. For insofar as it was an exercise of the will (and so an exercise of a practical power involved with deciding whether to *believe* that p) it will not have been a way of answering a theoretical question about the truth of a proposition, and thus not really a way of resolving the intractable theoretical deliberation for which the power was posited.

§IV. Conclusion

In this chapter, I have explored the claim that self-fulfilling beliefs, because of their exceptional structure, constitute an important exception to the widely accepted thesis that we cannot believe at will. Against this claim, I have argued that self-fulfilling beliefs are not an exception to the general rule. Believing at will is impossible and impermissible, even in those cases in which the belief in question is self-fulfilling. Understanding how the claim goes astray, however, requires a close examination of the nature of practical and theoretical rationality.

---

[36] An important point brought out by Hieronymi (2005).

[37] We can consider, for comparison, the ways in which having a power to form an intention at will, rather than pick an action at will, would circumvent, rather than provide resolution to, an agent deciding which bale of hay to pursue. (An important lesson from Kavka (1983)'s Toxin Puzzle case.)

I have begun by reconstructing Velleman's account of our special license to self-fulfilling beliefs. Velleman establishes this license by arguing for a form of epistemic permissivism and doxastic voluntarism with respect to self-fulfilling beliefs. Unlike other beliefs, self-fulfilling beliefs will be correct whether they are believed or disbelieved, and so either attitude – belief or disbelief – is, he claims, epistemically permissible. Since no one attitude is dictated by our epistemic reasons, we can choose what to believe at will on the basis of practical reasons.

I have also argued against this account, by considering each of these claims in turn. First, I have argued against the claim that in cases of self-fulfilling belief, we are permitted to believe or disbelieve the same proposition. I have distinguished between two possible readings of the aim of true belief: the aim of acquiring true beliefs, and the aim determining whether the believed proposition is true. I have shown that in cases of self-fulfilling belief, these two aims can come apart. I have argued that insofar as the aim of belief has to do with determining whether the believed proposition is true, Velleman's permissivism is implausible.

Second, I have argued that even if we accept that in cases of self-fulfilling belief our epistemic reasons suffice to adopt two different doxastic attitudes toward the same proposition, we cannot believe at will directly on the basis of practical reasons in such cases. If an agent is in a position where a belief that p and a belief that not-p would both be correct, she will need some mechanism to allow her to resolve her theoretical reasoning about p without recourse to further reasons, and this will preclude any mechanism which makes recourse to an agent's practical reasons for believing.

Chapter 4.1: Optimistic Reasons for Belief, Part I

Introduction

In the previous chapter, I argue against the possibility of epistemic permissivism — cases in which there are two different doxastic attitudes a believer might permissibly take toward some proposition. I argue that there can be no such permissivism, even when the truth of a proposition, p, depends on whether we believe that p.

This leaves us with a problem if, as I argue in chapter 2, our evidence under-determines any one doxastic attitude to take toward p in those situations where the truth of p depends on whether we believe that p. What attitude ought we to take with respect to such propositions, and on the basis of what reasons, if not our evidence, ought we to hold it?

In this chapter, I want to defend the Jamesian thesis that, sometimes, we ought to believe a proposition because it would be good if it were true. Many philosophers have been skeptical of such a claim because it seems to require accepting the existence of pragmatic justification for belief. That is, it seems to require accepting that our beliefs can be justified in virtue of considerations that show the belief useful, rather than true. I am also skeptical of such pragmatic justification. Rather than argue for the existence of pragmatic reasons for belief, I argue that in the relevant cases, failing to believe what would be best, if true, is an *epistemic* failing.

My argument will proceed in two parts. In part I, I will begin by distinguishing pragmatic reasons, which show something good about having a belief, from optimistic reasons, which show something good about the believed state of affairs. Much of the discussion of James has focused on the first sort of reason. This has led to a emphasis, in thinking about the permissibility of such beliefs, on issues involving the state/object distinction; pragmatic encroachment; and doxastic voluntarism – all of which are besides the point once we consider the permissibility of believing for optimistic reasons, instead. I will then consider the case against the possibility of Jamesian reasons,

so interpreted, and consider what conditions would have to be met for the possibility of believing for optimistic reasons. In the second part of this chapter, I will argue that these conditions can be met, and that believing for optimistic reasons is both possible and epistemically permissible.

§I. Pragmatic and Optimistic Reasons

A. An Ambiguity

Here is what James says about self-fulfilling belief:

> Often enough our faith beforehand in an uncertified result *is the only thing that makes the result come true.* Suppose, for instance, that you are climbing a mountain, and have worked yourself into a position from which the only escape is by a terrible leap. Have faith that you can successfully make it, and your feet are nerved to its accomplishment. But mistrust yourself... and you will hesitate so long that, at last, all unstrung and trembling, and launching yourself in a moment of despair, you roll in the abyss. In such a case (and it belongs to an enormous class), the part of wisdom as well as of courage is to *believe what is in line with your needs.*[1]

The position James endorses for self-fulfilling beliefs, as I will understand it, is that when engaged in theoretical reasoning with respect to some proposition, you ought to believe whatever is in line with your needs, or less self-interestedly, whatever is best.[2] Call those considerations which show a belief is in line with your needs "Jamesian reasons" for belief.

 But the term 'belief' suffers a state/object ambiguity: it can refer to *what* is believed – a proposition, or it can refer to the activity or state of believing that proposition – a propositional attitude. So too, the injunction to "believe whatever is in line with your needs" is ambiguous between at least two interpretations. The first is that you ought to adopt whatever *attitude* is best in line with your needs, the second is that you ought to believe in whatever represented *state of affairs* is best in line with your needs. Accordingly, there are two corresponding classes of reasons one might

---

[1] James (1985), 53.

[2] A full formulation (whether pragmatic or evidential) need to exclude cases where an agent has never considered a proposition, and so is not obligated to take any doxastic attitude toward that proposition, whatever their available reasons (c/f Feldman (2000)). Since I am dealing with those cases in which an agent is already actively engaged in theoretical reasoning about some proposition, it should not matter how those other cases are resolved.

refer to as "Jamesian reasons" to believe.[3] The first are those considerations which show something good about having the attitude or mental state, the second are those considerations which show something good about the believed state of affairs.

On the first interpretation of 'believe whatever is in line with your needs,' what matters for believing is not the content of your belief, but whether being in that belief state would be to your advantage. In James' example, if having the belief that you will fall to your death would allow you to jump the chasm, then believing that you will fall to your death would be believing what is best in line with your needs, because it would be the having of that psychological state which would lead to the best outcome.

This first interpretation is the interpretation Blaise Pascal appears to have in mind in his famous pragmatic argument for religious belief.[4] Since the potential benefits of believing that God exists (e.g. an eternity in heaven) are so great and the potential downsides so few, and since the potential downsides of believing that God does not exist (e.g. an eternity in hell) are so great and the potential upsides so few, there are strong reasons showing it to your benefit to have a belief that God exists, irregardless of the truth of your belief.

As the example of Pascal's Wager brings out, "Jamesian reasons" to believe that p in this first sense are closely related to one's practical reasons to *bring it about* that one believe that p.[5] Considerations which show something good about your being in some psychological state of believing that p are those considerations which count in favor of your taking action to get yourself to believe that p (regardless of the reasons, if any, for which you eventually end up believing it). If

---

[3] In this section, and throughout, I will take reasons to be considerations which bear on, or count in favor of, some act or attitude. Nothing rests on this assumption, however. For those who think of reasons as attitudes, one could replace the considerations I discuss with the contents of those attitudes.

[4] Pascal (1670).

[5] See Kavka (1989) and Hieronymi (2005) and (2009).

you can go to a philosopher who will convince you with argument that God exists, or if you can go

to a hypnotist, or take a pill which would directly implant the belief that God exists into your psyche,

then those considerations which show having the belief that p good to have will show you have

reason to listen to the philosopher or take the pill or visit the hypnotist. Believing for "Jamesian

reasons" in this first sense is thus also closely associated with *believing at will.* One will believe on the

basis of the sort of considerations which would count in favor of intending to believe, were

believing volitional. I will call the class of considerations which show something good about having

the belief that p 'pragmatic reasons' to believe.[6]

  Believing for pragmatic reasons is also an example of believing for what are sometimes

known as *state-given,* as opposed to *object-given* reasons for belief.[7] *Object-given* reasons, according to

Parfit, are the reasons we have for some attitude, "provided by facts about [the attitude's] object."[8]

Object-given reasons to desire, for instance, may include those reasons to desire some state of affairs

provided by facts about the desirability of that state of affairs. Object-given reasons to believe some

proposition may include those reasons to believe the proposition provided by the evidence for the

truth of the proposition. *State-given* reasons, by contrast, are those reasons which are "provided by

facts… about our *having* [the attitude]."[9] State-given reasons to desire, for instance, may include

---

[6] I use 'show something good' here loosely, in the same way we speak of reasons to intend an action as showing something good about an action. But like reasons for intending, practical reasons for belief need not not actually show something good about the belief. You need to take the considerations to show something good, but those might be bad reasons. You might be wrong about what shows an action good, or how good they show the action to be. Second, you need not even accept this 'guise of the good claim,' Insofar as it is possible for you to intend to act for the reasons which you take to show the action bad, we can count as pragmatic reasons to believe, those considerations which show the belief bad to have.

[7] See for example Parfit (2001) and Piller (2001)

[8] Parfit (1997) (22-23). In these passages, Parfit is speaking specifically about reasons for desiring. But he makes clear that he intends his discussion to apply to reasons for all sorts of attitudes (and activities), including including "reasons for believing, for caring, and for acting." (17)

[9] Ibid, 22-23.

those reasons to desire some state of affairs provided by facts about the desirability of *desiring* the state of affairs.[10]

Pragmatic reasons for believing are paradigmatic examples of state-given reasons for belief. The fact that you will be happier if you believe that there is an afterlife, for example, does not bear a rational connection to facts about the afterlife, but rather to facts about your psychology. This distinction is thus sometimes thought to mark off the divide between the right and wrong kinds of reasons for believing. Whereas right kind of reasons for believing (e.g. evidence) bear a rational connection to the object of belief by showing it true or likely, pragmatic reasons bear a rational connection to the state of believing, by showing the state of believing good or desirable.[11]

In contrast, on the second interpretation, what matters for the Jamesian position is not the benefit or utility of your being in some psychological state, but rather those good features of the believed state of affairs. Believing that p will be believing what is best in line with your needs just in case it would be best for you if p were so (whether or not you believe it).

This is the interpretation Bernard Williams has in mind in the case of a parent who, faced with evidence that their child was lost at sea, "very much wants to believe that his son is alive." Unlike Pascal's pragmatic believer, we can easily imagine a parent making such a claim where their desire would not be satisfied by visiting the local hypnotist who can bring it about that the parent acquires the relevant belief. They will not be satisfied, notes Williams, because "there is one sense - I think the more plausible one - of 'he wants to believe that his son is alive' in which this means *he*

---

[10] Though the reasons given by this fact may be *object*-given reason for the second order desire for the desire of some state of affairs. As Parfit observes, an important consequence of this point is that state-given reasons for one attitude may be object-given reasons for another. More on this in section II.C.

[11] Jamesian reasons in this first sense will also be closely connected to the issue of pragmatic encroachment - the possibility that contextual differences in the stakes of some subjects' belief can make a difference to the standard of evidence or justification necessary for a beliefs to count as knowledge (or to count as permissible).

wants his son to be alive - what he essentially wants is the *truth* of his belief." Here *believing what he wants* means believing what he *wants to be the case.*

Unlike believing for pragmatic reasons, believing for "Jamesian Reasons" in this second sense involves believing for considerations which we typically associate, not with volitional believing, but with *wishful thinking.*[12] Considerations which show something good about p are reasons to wish, or hope, or desire that p, not reasons to intend to believe, or bring it about that you believe, that p. I will call the class of considerations which show something good about p *optimistic reasons* to believe that p.[13]

Also unlike pragmatic reasons, believing for optimistic reasons involves believing for object-given, rather than state-given reasons. Optimistic reasons show something good about *p,* not the belief that p. And so they cannot be distinguished from evidence by the fact that they bear on the psychological state of believing that p, rather than on the believed proposition. Instead, optimistic reasons must somehow be distinguished from evidence by *the way* in which they bear on the believed proposition. Roughly, whereas evidence bears some rational connections to *the truth* of its propositional object, ones optimistic reasons for believing, like the standard object-given reasons for desiring or wishing that p, bear rational connections to *the goodness*, or the merits, of p (the subsequent section will be spent making this difference more precise).[14]

---

[12] This is the interpretation Rae Langton (2004) has in mind when speaking of "leaps of faith" or Grice (1971) when he complains of 'licensed wishful thinking.' Though there are also important differences between believing for optimistic reasoning, and wishful thinking, as it is normally understood (See section II).

[13] Like practical reasons, I use "show something good" here loosely, to parallel the way in which reasons for wishing are reasons which show something good about the wished-for object. But I want to allow that, insofar as there are bad or perverse reasons to wish for some object, there might be bad or perverse optimistic reasons for belief.

[14] Unlike Pragmatic reasons, Optimistic reasons are also independent from questions about contextualism, or pragmatic encroachment. While its true that how important some state of affairs is will vary from believer to believer, it will not vary from belief context to belief context, for a believer, nor will it match up with how important, in context, a belief might be.

Here, facts about the relative importance of her *belief* determines the appropriateness of knowledge attribution. Notice that it does not seem to matter how important the *subject matter* of the belief is. It might matter a great deal

B. *Separability*

Pragmatic and Optimistic reasons are not always carefully distinguished.[15] This is in part, I think, because when a belief is self-fulfilling both interpretations of the Jamesian position – the pragmatic and the optimistic – will usually be satisfied together. Notice that, were the beliefs of Williams' grieving father somehow self-fulfilling, his desire *would* be satisfied if he goes to the hypnotist: for in acquiring the self-fulfilling belief that p, he would have also made it the case that p.

More generally, if there is something good about some state of affairs, then there will be something good about the self-fulfilling belief in such a state of affairs – namely, that it will produce the good state of affairs. Thus, in the case of the chasm-jumper, the fact that you need to get to the other side of the chasm is both an optimistic reason to believe, in that it shows something good about jumping the chasm, and a pragmatic reason, in that it shows something good about having (or getting yourself to have) the belief that you will jump the chasm, insofar as having that belief will make it the case that you will jump the chasm.

Still, with a little ingenuity (and science fiction) we can imagine cases of self-fulfilling belief where the two reasons come apart. We can do so by adding incentive to hold the self-fulfilling belief in the sub-optimal state of affairs. Suppose your belief that you will get the hiccups is self-fulfilling. The fact that hiccups are painful shows something bad about having the hiccups, and so counts as an optimistic reason against believing that you will get the hiccups. Since your belief that you will get

---

whether there is life on other planets, though it may not matter at all whether you believe there to be life on other planets, and so such topics may have a no special standard of evidence. And it may matter very little which horse will win the race, though the stakes of your believing correctly may be very high, and so require a higher standard of evidence.

[15] A notable exception is Thomas Kelly (2003:182): "Wishful thinking should be carefully distinguished from the (alleged) possibility of basing one's beliefs on practical considerations. To a first approximation: wishful thinking involves holding a belief because one thinks that things would be better *if that belief were true*. This, of course, is quite different from holding a belief because one thinks that things would be better *if one held that belief*."

the hiccups will bring about your getting the hiccups, the fact that hiccups are painful also counts as a pragmatic reason against believing that you will get the hiccups.

But suppose now, too, that the eccentric billionaire from earlier chapters returns and offers you one million dollars to believe that you will get the hiccups. The fact that you will have a million dollars if you believe you will get the hiccups does not show anything good about having the hiccups. But it *does* count in favor of your believing that you will get the hiccups. Assuming the million dollars is worth the discomfort of the hiccups, we are now in a situation where, according to the pragmatic interpretation of the Jamesian position, you ought to believe that you will get the hiccups and where, according to the optimistic interpretation of the Jamesian position, you ought to disbelieve that you will get the hiccups.

If David Velleman's cognitivist account of intention is correct,[16] on which intentions are self-fulfilling beliefs about your actions, then Gregory Kavka's famous toxin case will be an example of a case with exactly this form. In the toxin case, the eccentric billionaire offers you money to adopt the self-fulfilling belief that you will drink some unpleasant, though non-lethal, toxin. Since the money is for merely adopting the attitude, not for drinking the toxin, it shows something good about having the attitude but nothing good about the object of the attitude – your drinking the toxin. One has pragmatic, but not optimistic, reason to believe that you will drink the toxin.

C. Belief and the "Counting in Favor of" Relation

With the two kinds of reasons clearly distinguished, we can now ask which of them might be the more plausible candidate to justify self-fulfilling belief. One reason why many have gravitated toward pragmatic reasons for belief has to do with a popular way of thinking about reasons in general, on which considerations become reasons for some attitude by counting in favor of that attitude. And though neither pragmatic nor optimistic reasons appear to count in favor of a belief

---

[16] For a summary, see chapter one, section III. For a more detailed account, see Velleman (1989) and Velleman (2000).

the way evidence does (by showing the believed proposition true) we have already seen another way in which pragmatic reasons might count in favor of a belief. They might count in favor *pragmatically*, or practically, for believing a certain way, in just the same manner that a consideration might count in favor of acting or intending to act a certain way.

Optimistic reasons don't even do that. In fact, as Thomas Kelly has observed, "perhaps the most objectionable thing about engaging in wishful thinking is that doing so often leads to bad consequences, in foreseeable and predictable ways."[17] Insofar as there is some general practical value to true beliefs for helping you navigate the world, wishful thinking will hinder your navigation, by systematically misinforming you of how the world is. You won't devote as many resources to studying or hunting or saving as you needed to if you believe, erroneously, that everything will work out for the best.[18] Thus, it might be hard to see how optimistic reasons are reasons for belief at all, let alone the right kind of reason to justify a belief.

It turns out to be notoriously difficult, however, to understand how the pragmatic way of 'counting in favor' of a belief by showing a belief to be good relates to the standard way in which evidence 'counts in favor' of a belief by showing the believed proposition true.[19] There are at best two plausible strategies for relating the two. First, one might try to understand our epistemic justification in terms of some practical aim (e.g. the importance of true beliefs for successful action), so that our epistemic justification for belief is a species of a larger genus of pragmatic justification. Second, one might try to understand how our believing can be governed by two different aims, such

---

[17] Kelly (2000): 182.

[18] For an empirical study of the consequences of wishful thinking, see e.g. Lang (2015). Though the practical implications of over- and under-estimation are actually far less clear than one might have thought.

[19] This way that evidence "counts in favor" of a belief is already a *very* specialized use of 'counting in favor.' It suggests, I think, some of the problems with thinking of reasons in terms of the 'counting in favor' relation once we move from reasons for action, to reasons for other attitudes. See Hieronymi (2005) and (2006) for further discussion.

that pragmatic reasons provide a distinctive 'pragmatic justification' for belief consistent with their epistemic justification.

Both strategies have been attempted with little success. We are not able to understand our epistemic norms in terms of practical aims. Our notions of epistemic warrant and justification stubbornly resist reduction to evolutionary norms or the norms of decision theory.[20] Nor, as recent work on the aim of belief has established, does there appear to be much space for non-epistemic aims governing belief. At the very least, either such strategy would require a radical departure from traditional thinking about the foundations of theoretical rationality, as grounded in truth, warrant, or knowledge.[21]

While good reasons for acting count in favor of the action, good reasons for believing show the belief true. We can say that such reasons 'count in favor' of the belief in a special way, but this counting in favor bears only a nominal connection to the way that practical reasons count in favor of actions. I think the fact that pragmatic reasons 'count in favor' practically of belief is a red-herring. It has led to an over-focus, in discussions of Jamesian reasons, on the relationship between our reasons for belief and the practical utility of a belief.

Once the temptation to think of pragmatic reasons as reasons for belief by courtesy – by virtue of some shared, attenuated, counting in favor relation – is resisted, there is no reason to think of optimistic reasons as any less likely candidates to count as reasons for belief. In either case, we would need a story of how Jamesian reasons for belief might fit into the standard picture of reasons for belief where a consideration counts as a reason for believing p by bearing on the truth of p.

---

[20] I cannot hope to embark in a full defense of this claim here, but I take it to be well-established by a variety of arguments in the recent literature on the topic. See, in particular, Burge (2009).

[21] I provide a more detailed argument of my own against such strategies in chapter 3.

And if this is our project, optimistic reasons turn out to be much better candidates than pragmatic reasons. For unlike pragmatic reasons, there *is* such a story that can be told for how it might be possible and rational to believe for optimistic reasons. Or so I will argue.

§II. Believing for Reasons

Before arguing that it is possible to believe for optimistic reasons, in this and the subsequent section I will first consider the question of why the skeptic thinks believing for Jamesian reasons – pragmatic or optimistic – is supposed to be impossible in the first place. Doing so will clear up the question of what it would be to believe for optimistic reasons, and what conditions would be required for believing for optimistic reasons to be possible. Having made clear the conditions for the possibility of believing for optimistic reasons, in the second part of the chapter I will argue that those conditions can be met.

Rather than working "from the top down," assuming we cannot believe for optimistic reasons, and then attempting to specify the conditions for believing for reasons in virtue of which it is impossible to believe for optimistic reasons, I will attempt to work backward from the other direction. In this section, I will look at cases where we do plausibly appear to believe that p on the basis of considerations which show p good, and use these cases to hone in on the sorts of cases the skeptic takes to be impossible, and so refine the notion of believing for reasons, according to which believing for optimistic reasons appears impossible.

A. Wishful thinking

One reason the notion of 'believing for reasons' needs refining is because, on the face of it, believing for optimistic reasons seems not just possible, but actual: we can find a large class of examples where we do seem to believe something on the basis of considerations which show it would be good if it were true in cases of *wishful thinking.* Consider, for example, the loyal fan who, in the face of mounting evidence to the contrary, persists in her belief that the Red Sox will win the

pennant, or the hopeful child whose belief that there will be a snow day persists into the morning despite seeing the roads cleared of snow.

In such cases, our beliefs appear sensitive not to our evidence (which would be insufficient for and perhaps count decisively against believing) but rather to the considerations for which we might *wish* the Red Sox to win, or the considerations for which we might hope for or desire a snowday. That is, our beliefs appear sensitive to those considerations which show there to be something good about what we believe, rather than those considerations which show what we believe is either true or likely.[22]

Do wishful thinking cases, where it is possible (if irrational) for us to believe something because it would be good if it were true, show that it is possible to believe for optimistic reason after all? They do not. At least, they do not amount to an example of the sorts of cases a skeptic about believing for optimistic reasons has in mind to deny.[23] This is because while cases of wishful thinking are possible, they are possible only in a very special range of circumstances. As Nishi Shah objects, wishful thinking fails to amount to a case of believing for optimistic reasons because "wishful thinking exerts its powerful effects only when masked. When I *consciously deliberate* about whether to believe that p, I cannot take the desirability of believing that p to be my reason for believing it."[24] Wishful thinking cases are those in which your beliefs about what would be good motivate your deliberation indirectly, operating either through self-deception or some non-reflective

---

[22] It's not obvious whether, when our wishful belief is motivated by an optimistic consideration, that consideration features as the content of another belief, or desire of the agents. It could be that the child's desire for a snowday explains her believing that it is a snow day, or it could be that the child's beliefs about all the fun she would have on a snowday explain her believing. I'm assuming the latter, though I think what follows will be compatible with either desire-based or belief-based pictures of moral psychology.

[23] Though some think it provides a limiting condition on a theory of *why* believing-for-optimistic reasons is impossible. See Shah (2013)

[24] Shah (2013): 315, emphasis added. See also Shah (2003) and Shah (2006). Shah falls prey, in this passage, to confusion between the difference between 'wishful thinking' on the one hand, involving the desirability of what you believe, and the desireability of believing, on the other.

psychological transition (e.g. the child who simply switches modes of representation, and 'slips' from imagining that there will be a snowday to believing that there will be a snowday).[25]

Wishful thinking can be understood as a special species of systematic deviant causation: in these cases, while our desire for p or beliefs about the good features of p *explain* our believing that p, the fact that our belief that p is explained by our beliefs about the good features of p does not yet mean that our beliefs about the good features of p are *the reasons for which* we believe that p.[26] Crucially, these sorts of wishful thinking cases are not those in which the positive considerations are employed as premises in your reasoning.

Contrast this with the sort of optimistic reasoning we have been considering in the millionaire case. This is a case in which a believer "consciously deliberates" about whether she will be a millionaire, and infers from the fact that her belief is self-fulfilling, and the fact(s) which show it would be good to be a millionaire, to the belief that she will be a millionaire. There is no easy analogous case of wishful thinking.[27] Indeed, it is misleading to speak of such cases of optimistic reasoning as 'licensed wishful thinking' at all. Unlike wishful thinking, these are cases of 'clear-eyed' optimistic reasoning. To find an analogue of such cases, we would need to find cases in which your believing that p on the basis of considerations which show p to be good is *directed* and *explicit*. It must be explicit in the sense that you can be aware that you are believing on the basis of wishful thinking,

---

[25] See, e.g., the Marx Brothers 'Shaking-Hands' scene in *Duck Soup* (https://www.youtube.com/watch?v=kcbglml9KRo) for a paradigmatic example of such a transition from supposition to belief.

[26] C/f Davidson (1963).

[27] Indeed, there is a powerful argument, from Williams (1973), that there could be no such case. For if one were to come to see that one's belief is motivated by wishful thinking, one would no longer see the content of that belief *as true* and so no longer count as believing (or at least, no longer count as believing *on the basis of* wishful thinking (c/f Winters (1979)) See Chapter 3, section IV for a longer discussion of this argument, in the context of self-fulfilling belief. While Setiya and Williams emphasizes the lack of awareness, it seems to me the lack of direction which is the more important point for the skeptic. I suspect it is possible to believe for reasons but be unaware of the reasons for which one believes.

and it must be *directed* in the sense that your thinking is *aimed* at resolving some inquiry, so that you can count as *inferring* p from your optimistic considerations.

B. Directed Reasoning

The phenomenon of wishful thinking, then, is not yet an example of believing for optimistic reasons. For while wishful thinking involves cases where a subject believes that p because she believes that p is good, it is not yet a case where a subject brings to bear her belief that p is good in her reasoning about whether p is so.

But there are also lots of cases in which considerations about the goodness of p *can* be brought to bear in an agent's reasoning about whether p is so. Consider the following case:

> Pangloss is engaged in theoretical deliberation about whether the plants in the garden have been watered. He knows that the plants need watering, and that Martin is in charge of watering the plants. Martin is conscientious, and does what he thinks is good. He is also practically wise, and so is usually correct in his judgments about what would be good to do. Pangloss thus believes that the plants have been watered, on the basis of the considerations which show that it would be good if the plants were watered.

In the above example, as in the cases of wishful thinking, we have a case where a subject's belief that the plants have been watered is to be explained by his belief in considerations which show it would be good if the plants were watered. Unlike cases of wishful thinking, however, Pangloss knows that his belief that the plants were watered is based on those considerations which show it good that the plants be watered. Pangloss' believing is *explicit*. More importantly, Pangloss's believing on the basis of optimistic considerations is *directed*. He is not only aware that his belief that p is good caused him to believe that p, his belief in the goodness of p caused him to believe that p because his belief in the goodness of p was part of his reasoning. He *inferred* that the plants were watered from the fact that it would be good if the plants were watered.

Pangloss is able to do this because in cases like his, and they are unexceptional, if not abundant, considerations which bear on the goodness of p do in fact bear on the truth of p, and so make perfectly good reasons for someone to believe that p. Unlike cases of wishful thinking,

Pangloss appears rational in believing that the plants were watered for the reason that it would be good if the plants were watered.

Someone might argue that, like cases of wishful thinking, these Pangloss cases will also be limited to a very special range of circumstances. But there are two important differences between the limits of wishful thinking and the limits of these Pangloss cases. First, unlike wishful thinking, what restrictions there are on Pangloss-cases are restrictions on what *propositions* we can rationally form beliefs about on the basis of optimistic considerations; they are not restrictions on the *circumstances* in which we can believe them. Pangloss cases, unlike cases of wishful thinking, do not appear to provide restrictions on *the manner in which* we might believe for optimistic reasons. There may be only certain propositions for which the fact of their goodness counts as evidence of their truth; but there are no restrictions, for those propositions, on when and how we can come to form beliefs about them on the basis of optimistic reasons.

Second, unlike cases of wishful thinking, even this restriction on the subject matter of potential Pangloss-cases seems like a contingent matter. Suppose, more radically, that we lived in some Leibnizian best-of-all-possible worlds. In such a world, considerations which showed it would be good if p were so would *always* be evidence for the truth of p, no matter what p involved. We can imagine physicists in such a world who, after being stumped with their attempts at proceeding via experimentation and the scientific method, instead try to determine how things are by thinking about how they ought to be (or, conversely, we might imagine normative ethicists in hard cases who, abandoning the process of a priori reasoning, begin tying people to trolley tracks to see who in fact gets hit.) In such imagined cases, there seems no limit to the kinds of states of affairs in which one might reasonably believe on the basis of considerations which bear on the goodness of those states of affairs. And so there seems no obstacle, in theory, to our coming to believe any state of affairs on the basis of the facts which show it would be good if it were so.

C. Employing Reasons

Such "Pangloss cases" show that we can sometimes believe a proposition, p, because we believe some consideration which bears on the goodness of p. And it shows that we can do this where that optimistic consideration is also *the reason for which* we believe that p. Nevertheless, I think these cases are also not yet sufficient to establish the possibility of believing for an optimistic reason, in the way the skeptic might have in mind to deny.

Recall, in our initial discussion of Jamesian reasons for belief, the similarity between pragmatic reasons for belief and believing at will, and optimistic reasons for belief and hoping or wishing. I propose, on behalf of the skeptic, that we take literally the connection between the two. To believe for Jamesian reasons just is to believe that p by willing to believe that p or by wishing or hoping that p.

What the skeptic of optimistic reasons for belief denies is not that the same sorts of considerations might both show p so and also show p good. Rather, what she denies is that one can believe that p by employing one's reasons in the way she would in wishing that p, and taking the reasons to show p good. Pangloss cases do not yet establish that this is possible, because there is a possibility that while some consideration, C is an optimistic reason for wishing that p, and while C is also the agent's reason for believing a proposition, the agent might, in treating it as evidence, still fail to employ it as an optimistic reason, and so fail to count as believing for an optimistic reason.

This is a possibility because reasons are fundamentally *relational*.[28] A reason is not just a consideration, but a consideration as it bears on some act or attitude. Thus the very same consideration might be two or more distinct reasons for several different attitudes. That it is raining, for instance, might be both a reason to believe one's flight is delayed, to hope no one left the windows open in the office, to fear the effects of global warming, to intend to carry an umbrella

---

[28] For the standard account, see Scanlon, (1998) ch1 and especially Scanlon (2014).

tomorrow, to believe that you ought to carry an umbrella tomorrow, etc. It is a reason to believe the your flight is delayed insofar as the fact that it is raining bears on the truth of whether your flight will be delayed; it is a reason to hope no one left the windows open in the office insofar as it shows something good or desirable about the windows being closed; it is a reason to intend to carry an umbrella insofar as it counts in favor of carrying an umbrella; and it is a reason to believe one ought to carry an umbrella tomorrow insofar as it bears on the truth of whether one ought to carry an umbrella.[29]

So too, the fact that I need to jump the chasm can count, in the appropriate circumstances, as several distinct reasons. It might be an optimistic reason to hope or wish that I will jump the chasm, insofar as it shows something good about jumping the chasm. It might also be a practical reason to get myself to believe that I will jump the chasm (if the belief is self-fulfilling), insofar as it shows something good about having the belief that I will jump the chasm.[30] The Plangloss cases show that, in some special circumstances, (if there is some divine being or other mechanism ensuring the co-extension of the good and the true) the fact that I need to jump the chasm might also be an evidential reason to believe that *I will* jump the chasm, insofar as it shows it true, or likely, that I will jump the chasm.

Whether a consideration is an *agent's reason for some attitude* depends on what role a consideration is playing for an agent, among the various kinds of reasons for various attitudes that consideration might be, which depends on how that consideration is being employed by the agent.[31] If, for example, the agent is employing the fact that it is raining in her theoretical deliberation about

---

[29] The last of these, between reasons to believe one ought to ø, and reasons to ø (or intend to ø) is of particular philosophical importance.

[30] Recall from S1b that for most self-fulfilling beliefs, *any* consideration which is an optimistic reason to hope that p will also be an apt practical reason to get oneself to believe that p.

[31] Where the distinction here is roughly that between a normative reason, and the operative reason.

whether her flight is delayed, she may be treating it as a reason to believe that her flight is delayed; if she employs the same consideration in practical deliberation about what to do, she may be treating it as a reason for acting or intending to bring the umbrella; if she employs it in her theoretical deliberation about whether some action is good, she may be treating it as evidence for the normative belief that she ought to carry an umbrella.[32]

Three relevant points follow. First, an agent might believe that it is raining, yet fail to employ this consideration in some or all of these ways. Though the fact that it is raining counts in favor of bringing an umbrella, an agent might not recognize its relevance, fail to bring to bear this fact in her reasoning about what to do, and so not bring the umbrella for the reason that it is raining.[33] Second, the same agent might be employing a consideration as several different reasons for different attitudes at the same time. The fact that it is raining might be both her practical reason for carrying an umbrella and her theoretical reason for believing that her flight is delayed. Finally, and most importantly, the fact that it is both a practical reason for her, and a reason for which she believes, does not mean that she is believing it *for a practical reason*. This point is obvious in the rain case – even though the fact that it is raining is, for her, both a practical reason, and her reason for believing, it is not a practical reason for believing – but it holds equally well in the Pangloss case. Some consideration, C might, in the right circumstances, show both that p is good and that p is true, and so be an optimistic reason for which one hopes that p insofar as one takes it to bear on the goodness of p, and the reason for which one believes that p, insofar as one takes it to count as evidence for the truth of p. But this does not mean that one has thereby believed that p by taking C to show p good, and so come to believe that p on the basis of an optimistic reason.

---

[32] This way of thinking about an agent's reasons bears close connections to Anscombe's notion of answerability. To be a reason for acting is to be an answer to a special sort of 'why-question.'

[33] Conversely, an agent might take as relevant some consideration which is not, in fact, relevant. An agent's reason for acting might not be a genuine reason to act.

Since reasons are individuated by their relations, the fact that a consideration can be both a optimistic reason to believe and a reason for which a subject believes, does not thereby show that the subject has believed for optimistic reasons. To do this, the subject must believe that p, for the reason that C, by employing C in the same way she would when wishing that p. And, the skeptic might argue, by employing C as evidence for p, Pangloss has failed to meet this requirement.

I think that the skeptic's argument is important, and that understanding it will give us insight into what is required for believing for optimistic reasons, and why meeting the requirements would be so difficult. In the next section, I will attempt to make the skeptic's argument more precise by adopting a more precise framework for kind-individuating reasons which can show us why optimistic reasons may appear, necessarily, the wrong kind of reason for which an agent might believe. Then, in section four, I will show how nonetheless, even on such a framework, in the case of self-fulfilling belief, we can believe for optimistic reasons after all.

§III. The Problem with Believing for Optimistic Reasons

A. Reasons and Inquiry

What the Pangloss case brings out is that what makes some consideration an optimistic reason to believe (and more generally, what type-individuates the kind of reasons one has) depends not on the content of the consideration but on how the consideration is employed.[34] Even the fact that p is good, for instance, might be either an optimistic reason or an evidential reason, depending on whether it is being brought to bear on establishing the goodness or the truth of p.

To mark this link between the reasons for which an agent holds an attitude, and the manner in which a consideration is employed in reasoning, I will adopt a helpful way of thinking from Pamela Hieronymi, and treat a reason as a consideration that *bears on a question (or conclusion)*.

---

[34] A point going back at least as far as Frege (1997) and Geach (1965).

Hieronymi takes this formulation to capture the general observation that "a reason is an item in reasoning," where reasoning, in contrast to a mere series of thoughts or set of claims, is "directed" in that the reasoning functions to establish some conclusion or answer some question or another. Pieces of reasoning "bear 'rational relations' to one another and thereby bear on the conclusion [or question]" toward which the reasoning is directed.[35]

We can mark the different kinds of reasonings in which a consideration might be employed by the kind of question or conclusion the reasoning is directed toward. Practical reasoning, for instance, is reasoning directed toward questions of what to do. Theoretical reasoning is reasoning directed toward questions of what is so. Since reasons, thought of as possible items of some reasoning, are individuated by the kinds of reasoning in which they are possibly employed, and since kinds of reasonings are individuated by the kind of questions toward which they are directed, we can distinguish kinds of reasons by the kinds of questions on which they bear, or might be taken to bear.

What sort of reason a consideration is, *for an agent*, depends on what question, or conclusion, that agent is taking that consideration to bear on.[36] The agent might mistakenly take some consideration to bear on a question, or give it inappropriate weight or valence. It is thus possible

---

[35] Hieronymi, (2005) 443-444. The reasoning here is thus directed at a question in a relatively weak sense. The agent needn't explicitly see herself as reasoning in order to answer some question, or formulate the question to herself explicitly beforehand.

[36] There is an enormous controversy and disagreement in the analysis of what constitutes a taking a reason to bear, or the 'basing-relation,' and it may seem that this account involves a contentious commitment to a strong answer. I take this talk of questions to be broadly ecumenical, in that it is a formal account, compatible with many different substantive psychological or moral-psychological accounts of what bringing a consideration to bear in answering a question amounts to. (e.g. in terms of counter-factuals, capacities, dispositions, intuitions, rational causal transitions, explanations, etc.) But I also think the question of whether we can believe for pragmatic reasons is best adjudicated at this level of description, and that the arguments here will be compatible with any account of what it is to employ a consideration in one's reasoning.

that some consideration be an agent's reason for concluding some question, though the consideration does not genuinely bear on that question.[37]

These reasons become reasons for attitudes in virtue of constitutive connections between certain propositional attitudes, and the settling of certain kinds of questions.[38] Thinking of reasons for attitudes by way of reasons bearing on questions captures two important and closely-related features of many reasons-sensitive propositional attitudes.

The first feature involves a descriptive claim, marking the relationship between certain kinds of attitudes and certain kinds of reasoning.[39] Certain attitudes are the kinds of attitudes they are because they are the kind of attitude which are either (partly) constituted by, or regularly issues from, the positive conclusion of certain kinds of deliberations, inquiries, or directed-processes.[40] Part of what it is for my attitude to be a *belief* that p, as opposed to some other attitude like an imagining that p or a desire for p, is that it be the kind of state which issues from (or is partly constituted by) the conclusion to a piece of *theoretical* reasoning about whether p is so.[41] Whereas part of what it is for my attitude to be an *intention* to ϕ , is that it be the kind of state which issues from (or is partly constituted by) the conclusion to a piece of practical reasoning about what to do.

While not all beliefs and intentions need issue from some explicit deliberation, the fact that they are the kind of attitude which issued from *some* process which functioned to track the truth, or

---

[37] Further, there might be circumstances in which the correct answer to some question is 'no,' but that the agent, given the considerations she has available, would be reasonable in answering 'yes.'

[38] Ibid, 449.

[39] Ibid,447; 449.

[40] This is a broadly functional account. I say partly constituted by their processes that produced them, because they are also functionally individuated by their 'downstream' rational connections to other parts of our mental lives.

[41] I assume here some broadly functional account of mental states. I say partly constituted by their processes that produced them, because they are also functionally individuated by their 'downstream' rational connections to other parts of our mental lives.

determine what to do, respectively, is part of what explains why they play the distinctive functional role in our mental architecture that they do.

The second feature involves a normative claim. It is constitutive of certain attitudes that they embody certain kinds of rational commitments, or resolutions, expressible as the settling of certain types of questions.[42] This comes from an influential way of thinking about justification, from Elizabeth Anscombe, in terms of *answerability*.[43] Anscombe observes that part of what it is to act intentionally just is for you to be responsible to a certain sort of request for answer to the question 'why are you doing that?' where your answer to that question gives your reasons (if any) for acting. Your reasons for acting are those considerations which justify your acting by providing the considerations in light of which you have concluded acting as you are is the thing to do. So too, part of what it is to believe that p is for you to be responsible to a certain sort of request for an answer to the question of whether p is so.[44] Your reasons for believing are those considerations which justify your believing by providing those considerations in light of which you have concluded that p is so.[45]

These two points are related. What you are answerable to, when you are responsible for a certain attitude, is the question the answering of which is constitutive of, or regularly issues in, that attitude.

B. *Wishing, Willling, Believing*

We can distinguish reasons for various types of attitudes by distinguishing the questions, a commitment to an answer to which is constitutive of (or regularly issues in) that type of attitude.

---

[42] Ibid, 450.

[43] For a sampling of the rich discussion of the nature of this answerability, see Anscombe (1957), Moran (2001), and Hieronymi (2009).

[44] It is a substantive claim that the relevant question for which a believer is answerable is the first order 'whether p?' rather than the second order 'why believe that p?' (see Velleman/Shah (2003)). I argue for this claim in chapter 3.

[45] Hieronymi (2005) 450.

The fact that beliefs aim at the truth, for instance, is to be understood as the fact that a belief that p embodies a rational commitment to the truth of p, expressible as a commitment to an affirmative answer to the question of whether p. A belief that p is correct just in case p really is so, and so the correct answer to the question of whether p is an affirmative one.

Intending, in contrast, is not associated with questions of what is so, but rather with questions of what to do. An intention to φ embodies a rational commitment to φ-ing as the thing to do, expressible as an affirmative answer to the question of whether to φ. An intention to φ is correct just in case φ really is the thing to do, and so the correct answer to the question whether to φ is an affirmative one.[46] This commitment is a commitment not just to the truth of the claim that φ-ing is choice-worthy, but also a practical commitment *to acting,* and thus distinct from the belief that one ought to φ.[47]

Like intending, wishing that p (or at least an important subclass of wishing that p) is distinct from a belief in the goodness of p in that it involves not just a commitment to the truth of the claim that it would be good if p were so but also a practical commitment to your having it be (if you could) that p.[48] While such wishing shares with intending an action-like commitment, wishing is unlike intending in that we can wish for a broader class of objects than we can intend, because we can wish for things outside of our power to do or bring about.[49] Wishing, then, is roughly equivalent

---

[46] See Appendix 'On the Aim of Intention' for further discussion.

[47] This formulation is intended to capture the Aristotelian idea that practical reason concludes in action.

[48] I take this to be at least an important subclass of wishing, though I leave open the possibility of other senses of 'wishing' further divorced from such practical commitments to action (i.e. the sense in which we might intend undertake some unpleasant task, but wish that something prevents us.)

[49] This account of wishing owes a great deal to Aristotle's account of wishing in NE 3:iii. I argue for the practical nature of such wishing in more detail in chapter 5.

to "willing off-line."[50] If willing involves reasoning about what you will have be done, this wishing involves reasoning about what you *would* have be done. There is some practical question, the answering of which is constitutive of a wish that p, rather than a belief about what should be so; just as the answering of the practical question whether to φ, the answer of which is constitutive of intending to φ, is different than answering a merely theoretical question of what one ought to do. I will mark this distinction by describing the question, the answer to which is constitutive of wishing, as the question: 'would that p?'

C. The Wrong Kind of Reasons

Recall that what the skeptic of Jamesian reasons for belief denies is not that the same sorts of considerations might both show p true and show p choiceworthy or good. Rather, what she denies is that one can believe that p by employing one's reasons in the way one would in wishing that p or intending to believe that p. Thinking about reasons in terms of questions on which they bear allows us to see why this might be impossible.

The skeptic does not deny that the same considerations might bear both on the question of whether p is so and on either the question of whether to believe that p or the question 'would that p.' Instead, for the skeptic, believing for Jamesian reasons amounts to believing that p by taking those reasons to positively settle the pragmatic question of 'whether to believe that p' or the practical question of whether p is to be so (or whether you would have p be so.) Believing for optimistic reasons would involve the latter.

But since what it is to believe that p is to be committed to a positive answer to the question of whether p is so, if you 'believe' that p by taking some considerations to settle the question of whether you would have p be so, the attitude you have constituted would not really be a belief after

all. For you did not take your considerations to establish a positive answer to the question 'whether p.' You have taken them to establish a positive answer to a different question: the question 'would that p.' You would thus have arrived at a *wish* that p. Wishing, after all, just is the attitude regularly issuing from, or embodying, a commitment to a positive answer to the question 'would that p.'[51]

Even if you are also committed to a positive answer to the question of whether p is so, and even if (as in the Pangloss case) you are also committed to a positive answer to that question by the very same considerations for which you are committed to a positive answer to the question 'would that p,' you would not have come to believe that p *by* becoming committed to a positive answer to the question 'would that p,' and so you would not have believed by employing those considerations as optimistic reasons.

The same sort of argument rules out the possibility of pragmatic reasons for belief. To believe for pragmatic reasons would be to believe by taking some consideration or set of considerations to settle the question of whether to believe that p. But if you 'believe' that p by taking some considerations to settle the pragmatic question of whether to believe that p, you would not have arrived at a belief, because you would not have settled the question of whether p is so. You would instead have arrived at an intention, or decision, to believe or get yourself to believe.

§IV. Possible Cases of Believing for Optimistic Reasons

We are now finally in a position to state more precisely what, according to the skeptic, would be required for a subject to believe for optimistic reasons, and to see why it is that the skeptic takes it to be *impossible* that an agent do so. For a consideration, C to be an agent's optimistic reason is for the agent to be employing C in her practical reasoning, by bringing to bear the consideration in settling a practical question of whether she will (or would) have p be so. For a consideration, C, to be an agent's reason for her belief that p is for the agent to bring to bear that consideration in

---

[51] See Hieronymi (2005) and Hieronymi (2009).

settling the theoretical question of whether p *is* so. For an agent to believe for an optimistic reason, then, would be for an agent to come to believe that p by bringing C to bear in settling the practical question of whether she will (or would) have p be so. But if she does this, she has not believed, but wished; since to believe she must instead have settled the question of whether p is so.

But this argument for the impossibility of believing for optimistic reasons is too quick. For notice that the fact that believing that p for an optimistic reason requires believing that p in virtue of answering the question 'would that p' does not (on its own) show that believing that p for an optimistic reason is impossible. What makes an attitude incompatible with being a belief that p is that the subject of the attitude fail to have the attitude in virtue of her commitment to the truth of p, constitutive of her belief that p. Belief is ruled out, not because a subject has that attitude in virtue of a commitment to an answer to some practical question, but because the subject *fails* to have the attitude in virtue of a commitment to an answer to a theoretical question of whether p is so.

The skeptic's argument against believing for Jamesian reasons thus requires that we also accept the further premise: that it is impossible, in answering a practical question like 'would that p,' to thereby commit oneself to the truth of p. If, instead, having an attitude in virtue of answering the question 'would that p?' were compatible with having that attitude in virtue of committing to the truth of p, it might yet be possible to believe for optimistic reasons.

Of course, a premise denying the possibility of committing oneself to the truth of p in answering the question 'would that p' may not seem so implausible. Since the question 'would that p' and the question 'whether p' are two different questions, it may seem that insofar as you are employing those considerations in answering the question 'would that p' you are not employing

them in answering the question 'whether p' and so not (thereby) committed to p being so.[52] And, as Hieronymi points out, it is hard to make sense of what it could mean to say that you can "commit yourself to the truth of p by finding convincing reasons that you do not take to show it true."[53]

Having an attitude in virtue of answering the question 'would that p' would be incompatible with having that attitude in virtue of a commitment to the truth of p, if we accept that the answering of questions is exclusive in the sense that insofar as we are answering one question we are not answering the other. If the answering of questions is exclusive, then insofar as we take our reasons to settle a practical question about what we would have be so, we are not taking them to settle the theoretical question of what is so, and so not taking them to show p true, and so not, in virtue of finding them convincing, committed to the truth of p.

If, on the other hand, we could find some way to bridge the questions 'whether p?' and 'would that p?' such that in answering one question, an agent was answering the other, we could explain how it was possible for that agent to believe for optimistic reasons.

The challenge for the optimist is to show how one class of reasons – reasons which show p is so – could be systematically connected to an apparently unconnected class of reasons, reasons which show p should be so. Thinking about reasons in terms of questions on which they bear shows why this is so difficult. To believe for an optimistic reason would require that we bring a consideration to bear, in the same piece of reasoning, in answering a theoretical question about what is so and a practical question of what to do or what you would have be so.

I believe, however, that there are two philosophically and important cases where this, plausibly, is exactly what happens. These cases are those surveyed in chapter 1: self-knowledge of

---

[52] Of course, you might *also* take those reasons to show p true. But not insofar as you are thinking them to answer the question 'would that p.' Even in the rare circumstance where some consideration happens to bear on both of these questions, one cannot bring it to bear on both at the same time, or in the same employment.

[53] Hieronymi (2009)

our mental states and practical knowledge of our actions. In the the rest of this section, I will argue that our capacity for self-knowledge and practical knowledge can both be fruitfully understood as a capacity to believe for optimistic reasons. In Part II of this chapter I will present a theory of how this could possibly be so, given the difficulties raised here in part I, by looking more closely at the case of practical knowledge.

A. Practical Knowledge

Consider first the case of practical knowledge knowledge of our actions. Practical knowledge involves those beliefs an agent has about her own actions; (i.e. beliefs in propositions of the form: I will φ, or I am φing, where φ is some possible action or activity). Practical knowledge is distinctive, in part, because of important first-person and third-personal asymmetries associated with beliefs about actions. Unlike my predictions about the actions of others, which is based on behavioral or psychological evidence, my beliefs about my own actions are non-observational. We appear to know what we are doing "immediately" without apparent recourse to the sorts of inferences others must use if they are to third-personally form the same belief about us.[54] Insofar as we do come to the belief inferentially, it involves a dramatically different pattern of inference. As Elizabeth Anscombe has observed, the sorts of considerations on the basis of which I believe that I will act appear to differ sharply from the sorts of considerations on the basis of which I believe others will act.

Consider, for example, the case of two coffee-drinkers. If I am asked whether you will have another cup of coffee, I may bring to bear statistical evidence, such as facts about your habits and history of coffee-drinking. I may bring to bear behavioral evidence, such as the fact that you have been staring longingly at the coffee pot, or gesturing toward the waiter. I may also bring to bear some 'non-observational' but psychological, evidence. I may, for instance, infer from the fact that

---

[54] See especially Anscombe (1957); Setiya (2003)

you desire to stay up late, and your belief that that coffee will keep you awake, to the conclusion that you will have another cup of coffee.

In contrast, when asked whether or not *I* will have another cup of coffee, I do not appear to need to consult general statistical evidence, my past history, or my psychology. In general, we need not infer our belief that we are acting from evidence – either observational facts about our behavior or psychological facts – to believe that we will perform some action. I can answer for myself the question of whether I will have another cup of coffee by thinking, not about my observable evidence, but instead by thinking about whether or not to have another cup of coffee.[55] I can conclude that I will have another cup of coffee on the basis of the fact that I need to stay up late, and the fact that coffee will keep me awake.[56][57] Unlike my predictions about the actions of others, says Anscombe, which are based on observation or on evidence, my predictions about my own actions are "non-observational," in the sense that at least in some cases, my beliefs "are justified, if at all, by a reason for acting, as opposed to a reason for thinking them true."[58]

Notice that, whether or not you agree with Anscombe's claim about our non-observational knowledge, the "reasons for acting" Anscombe mentions here, on the basis of which a person knows how she will act, are optimistic, not pragmatic, reasons for believing. When you believe that you will have another cup of coffee for the reason that coffee will keep you awake, you are bringing to bear, not those considerations which count in favor of having the belief that you will have that second cup of coffee, but rather those considerations which count in favor *of your having* the coffee.

---

[55] See Velleman (1989); Moran, (1999); Wilson (2000)

[56] Though, in my beliefs about whether you will drink coffee, I might take as a reason your *beliefs* about whether coffee will keep you awake, in the case of my beliefs about whether I will drink coffee, I take as a reason the fact about coffee itself – a fact which bears, not on my psychology, but on merits of coffee-drinking.

[57] This claim is contentious. For an alternate view, on which we infer what we do from our psychology, see Grice, (1971) and more recently, Sarah Paul (2009). I discuss issues with such 'inferrentialist' accounts in chapter 4.2.

[58] Anscombe (1957).

If Anscombe is right, then practical knowledge of our actions involves believing for optimistic reasons: we believe that we will act a certain way because we should act that way. Where p is of the form: I will φ or I am φing, we believe that p on the basis of considerations which show p good or choiceworthy.[59]

B. *Self-Knowledge*

Consider too the case of self-knowledge of our mental states. Self-knowledge involves those beliefs an agent has about their beliefs or mental states more generally (e.g. beliefs in propositions of the form: I believe that q). Like the case of practical knowledge, self-knowledge is distinctive, in part, because of the distinctive first-personal/third-personal asymmetries associated with beliefs about our mental states. There is a sharp difference, in many central cases of self-knowledge, between the sorts of considerations for which I believe that I believe, and the sorts of considerations for which I believe that you believe. When forming a second-order belief about whether you believe that q, I bring to bear those considerations which bear some rational connection to your psychology: facts about your other beliefs, your cognitive ability, etcetera. If, for instance, I were to form a belief about whether you believe that it is raining, I might consider whether you have been outside, the well-functioning of your perceptual systems, whether you are a frequent weather-channel watcher, etc.

In contrast, when forming a second-order belief about whether *I* believe that q, I do not survey facts about my psychology or behavior. I survey, instead, my evidence for q. My thinking about what I believe is *transparent to* the subject matter of those beliefs.

---

[59] I do not mean to imply any strong 'guise of the good' thesis here. All an agent need to is believe on the basis of considerations she takes to count in favor of p (in this case, of her φing). Insofar as perverse agents can take the fact that some act is bad to count in favor of doing it, such a fact would function as an optimistic reason.

The motivation for the transparency account comes from the work of Gareth Evans, who

observes that when we ascribe beliefs to ourselves, we survey, not facts about our psychology, but

facts which bear on the truth of the content of the belief we are ascribing. Evans observes:

> In making a self-ascription of belief, one's eye's are, so to speak, or occasionally literally, directed
> outward – upon the world. If someone asks me 'Do you think there is going to be a third world war?' I
> must attend, in answering him, to precisely the same outward phenomena as I would attend to if I
> were answering the question 'Will there be a third world war?' I get myself in a position to answer the
> question whether I believe that p by putting into operation whatever procedure I have for answering
> the question whether p.[60]

If I am asked whether I believe that it will rain tomorrow, for instance – a question about my state

of mind – I don't think about my psychology, but rather about the weather. More generally, when p

is of the form: I believe that q, I can form the belief that p on the basis of those considerations

which show q true.

But, as Nishi Shah has pointed out, the considerations which bear on the truth of q (e.g. my

evidence) are precisely those considerations in virtue of which I ought to believe that q. So to say

that we believe that we believe it is raining on the basis of considerations which show that it is

raining, is to say we believe that we believe that it is raining on the basis of those considerations

which show we *should* believe that it is raining.[61]

More generally, where p is of the form: I believe that q, we can believe that p on the basis,

not of considerations which show p true, but which show q true. And since those are the same

considerations for which we ought to believe q, you come to believe that you believe that q by

figuring out whether to believe that q in the same way that you come to believe that you will φ by

figuring out whether to φ.[62] As in the case of practical knowledge, when p is of the form (I believe

[60] Evans (1982).

[61] Shah (2003).

[62] I may appear to be adding a new use of 'transparency' to an already crowded list. Moran (2001) thinks of the relation of transparency as holding between the question: "do I believe that p?" and "p?", Shah (2003) thinks of the relation of transparency as holding between the question "whether to believe that p?" and "p?" and I am thinking of the relation as

that q) you are believing that p on the basis of those considerations which show the merits of p. You are believing for your optimistic reasons.[63]

This affords us an explanation of one of the more puzzling aspects of many transparency accounts of self-knowledge: the connection between the considerations you believe for and the proposition you are believing, a connection which has looked, at least on the surface, to be mysterious. As Moran poses the question:

> why, then… is it legitimate to answer the question [of what attitude one has] in a way that seems to neglect the fact that is about a particular person, and instead treat it as a question concerning the topic of the attitude itself?… it would not, in general, make sense to answer a question about my state of mind by attending to a logically independent subject matter.[64]

If we can believe for optimistic reasons, we have a potential answer to this puzzle. It is true that evidence for p is not evidence that we believe that p. But it is a mistake to think that there is *no* relationship between our evidence for q and our second-order belief that we believe that q. Because evidence for q constitutes a reason for which we should believe that q, it is an optimistic reason for believing that we believe that p. It shows the object of our self-knowledge to be good.

§V. Conclusion

If I am right, I have isolated two cases where, plausibly, agents appear to believe for optimistic reasons – they appear to believe that they believe or act a certain way on the basis of considerations which show that they ought to believe or act that way. Three things remain to be

---

holding between the question "Do I believe that p?" and "should I believe that p?" In fact, I think these three uses are all species of a more general phenomenon: A question, Q, is transparent to some other question, V, for some agent, S, if S can answer Q by answering V. A Belief or class of beliefs, has the property of transparency if the question, the answering of which amounts to the forming of that belief, is transparent to some other question. For example, the question: do I believe that p? (the answering of which amounts to the forming of the second-order belief that one believes that p) is transparent, for me, to the question: whether p? I will discuss this phenomenon in more detail in chapter 4.2.

[63] This point may appear to involve a certain amount of slippage. For while the considerations which show p true are considerations which show the belief that p good or valuable, they show the belief that p to be *epistemically* good or valuable. But these cases can be accounted for if we abandon entirely the talk about whether some state of affairs is 'in line with ones needs' and switch to a broader conception of whether a state of affairs is valuable.

[64] Moran (2012): 213.

shown. The first is that these are genuine cases of believing for optimistic reasons – that is, that

agents are not simply treating the fact that they ought to act or believe a certain way as evidence, as

in the Pangloss cases of section III. The second thing that remains to be shown is how such genuine

believing for optimistic reasons is possible, given the arguments surveyed in section IV. Supposing

that we can explain how believing for such optimistic reasons is possible, the third thing that

remains to be shown is that, in believing for optimistic reasons, an agent can be epistemically *justified*.

In the second part of this chapter, I will attempt to show all three.

Chapter 4.2: Optimistic Reasons for Belief, Part II

Introduction

In this chapter, I defend the thesis that, sometimes, we ought to believe whatever would be best, if true. Though believing what is best may appear to be a form of wishful thinking, and so unwarranted, I will develop an account according to which optimistic reasons which show what we believe to be good can also be genuine epistemic reasons for believing it to be true.

In Part I of this chapter, I distinguished pragmatic reasons, which show something good about having a belief, from optimistic reasons, which show something good about what we believe. In Part II I will argue that, unlike pragmatic reasons, optimistic reasons can be the proper grounds for a distinctive form of reliable, truth-directed, inference: determining what is true by deciding what is to be true. Since optimistic reasons, so employed, help a believer determine whether her belief is true, these optimistic reasons can be genuine epistemic reasons for that belief.

While I believe that this account holds the key to understanding how we are to reason when our beliefs are self-fulfilling, I will proceed by looking at a slightly different case: the case of practical knowledge of our actions. I will argue that when we are forming beliefs about states of affairs under our control, we enjoy a special way of engaging in theoretical deliberation about their content: rather than discovering what is true on the basis of evidence, we can instead decide what is to be true on the basis of optimistic reasons.

Applying this result, in the next chapter I will argue that self-fulfilling beliefs are best understood as a special instance of the larger phenomenon of our capacity to believe for optimistic reasons in cases where we have control over the states of affairs we are reasoning about.

§I. Belief, Intention, and Deliberation

One of the central challenges in the philosophy of action involves explaining the relationship between an agent's intention to perform an action and her belief that she will perform that action.[1] In chapter 4.2, I will survey the two general schools of thought about the relationship – the *cognitivist* position, on which we identify an agent's intention to act with her belief that she will act and the *inferrentialist* position, on which an agent infers the belief that she will act from her intention to act – and the competing pressures that have pushed philosophers toward each view.

I will then propose and argue for an alternative picture, one which I think can accommodate the competing pressures for both sides and one which also presents an independently attractive picture of the relationship between belief and intention. The position I want to argue for is a partial hybrid – rather than identifying the belief and the intention we should instead identify the *reasoning* by which we arrive at each: in the appropriate circumstances, the same reasoning can give rise to two distinct attitudes, an intention and belief, with the same propositional contents. When this occurs, we are able to believe that we will act (and justified in so believing) on the basis of reasons for acting.

My argument will proceed in three steps. I will begin by noting certain structural similarities among the pressures for and against identifying belief with intention. I will argue that one could resolve both sets of pressures if it were possible to have two distinct mental states which issue from one and the same piece of reasoning. In the second part of my argument, I will argue that it *is* possible to have two distinct mental states which issue from the same piece of reasoning. I will do so by addressing a potential problem with such a

---

[1] While I discuss the relation between beliefs and future intentions, much will also apply to the relationship between beliefs and intentions-in-acting. For some of the nuances involved in the relationship between these two sets of issues, see Falvey (2000).

possibility raised in chapter 4.1: that it requires a problematic blurring of the conceptual distinctions between practical and theoretical reasoning. I will conclude by addressing several objections to the effect that the picture I have sketched, once fully specified, fails to constitute a genuine third alternative. I will argue that this new view *does* manages to carve out a genuine middle ground between the cognitivist and inferrentialist view without collapsing into one or the other.

§II. Cognitivism and Inferrentialism

I will begin by briefly sketching the sorts of pressures which have pushed thinkers toward and away from identifying beliefs with intentions. That one might think the two mental states are tightly related has a variety of sources. One is the similarity of their representational contents, as marked in our surface-level grammar. "I will go for a walk," can equally well be an expression of the belief that one will go for a walk, or of one's intention to go for a walk.[2]

This surface-level grammar is mirrored by a deeper connection between an agents acting intentionally and knowing (or at least believing) that she is acting, under that intentional description. A marker of a certain action being un-intentional – say my humming aloud in the office – is that I did not realize I was doing it.[3] Indeed many have held that an agent's not believing that she is acting is a *sufficient condition* on her action being unintentional and, conversely, that it is a necessary condition on acting intentionally or intending to act that an agent believe she is, or will, so act.[4]

---

[2] Anscombe, (1957): §2

[3] Anscombe, (1957): §6.

[4] This is a point made much of in the cognitivist account of Setiya, in a series of works: Setiya (2003); Setiya (2007); Setiya (2008); Setiya (2011).

There are, of course, standard objections to the claim that intending and believing are always necessarily connected. We can consider cases of 'difficult action,' such as the example, from Davidson, in which an agent intends to make ten carbon copies but, because of the difficulty involved, does not believe that she will succeed. Or we can consider the example of the agent who intends to quit smoking, but, cognizant of the likelihood that she will succumb to inevitable temptation, fails to believe that she will successfully follow through.[5]

But there are also standard responses to these standard objections. Even in cases where you intend to act a certain way but do not believe that you will succeed, there is at least *some* more basic action you are performing intentionally or intending to perform, that you *do* believe that you are doing or will do, under that intentional description. I may not believe that I will quit smoking, but I believe that I will *try* to quit smoking. I may not believe that I will make ten carbon copies, but I believe that I will press down hard on the paper, and having *this* belief is a necessary condition on the whole action being intentional.[6]

In any case, it seems as though, at least in central cases, there is some very close connection between intending to act, and believing that I will or am so acting. This connection between my action being intentional and my believing that I am acting seems to be a connection in need of explaining, and one explanation would be in terms of a tight relationship between my intention and my belief.

---

[5] See, for example, Marusic (2012).

[6] An argument employed forcefully in Setiya (2008). There are also many other, increasingly sophisticated, responses to Davidson-style objections. Velleman (1989) argues that in cases where an agent foresees sufficient difficulty with φ-ing, one cannot intend to φ, but can only intend to try, or take steps toward, φ -ing. Schwenkler (2015): 14-16 and Thompson (2011): 210 have argued that in cases like that of the carbon-copy machine, your believing that you will φ follows from the *guiding* role of the intention to φ. If an agent intends to make ten carbon copies, she will believe that she will succeed, despite the difficulty involved, because she intends to continue working toward her goal until she eventually achieves it, even if she encounters initial failure.

A second set of pressures for closely identifying belief and intention stems from important first-person and third-person asymmetries associated with beliefs about actions.[7] One asymmetry concerns the special first-personal *authority* a believer often has with respect to beliefs about her own actions. In central cases, an agent is in a better position to know what she is doing than others are. If I believe that I am going to have another cup of coffee, for example, you ought not to bet against me.

A second asymmetry concerns the special first-personal *access* a believer has with respect to beliefs about her own actions. Unlike my predictions about the actions of others, which is based on behavioral or psychological evidence, my beliefs about my own actions are non-observational. We appear to know what we are doing "immediately" without apparent recourse to the sorts of inferences others must use if they are to third-personally form the same belief about us.[8] Insofar as we do come to the belief inferentially, it involves a dramatically different pattern of inference. The sorts of considerations on the basis of which I believe that I will act appear to differ sharply from the sorts of considerations on the basis of which I believe others will act.

Following Anscombe, many have thus argued further that we seem enjoy a special sort of *practical knowledge* with respect to beliefs about or own actions: my beliefs about my own actions are "justified, if at all, by a reason for acting, as opposed to a reason for thinking them true."[9] Since these are the very same considerations for which I come to intend to act,

---

[7] A point whose importance is highlighted by Anscombe (1957) and Hampshire (1959). For more recent development of this line of thought, see Velleman (1989); Moran (2001); and Wilson (2000).

[8] See especially Anscombe (1957); Setiya (2003)

[9] Anscombe (1957): 6. Anscombe is in turn following Thomas Aquinas' (ST IaIIae q3, a.5) conception of practical knowledge as "the cause of that which it understands."

the puzzling phenomenon of practical knowledge is also to be potentially explained by some important relationship between my intention and my belief.[10]

Of course here too, there is no consensus about the full extent of the asymmetry. In particular, not everyone has been convinced by the phenomenon of practical knowledge. For many, this is because it does not seem possible to *believe*, and certainly not possible for that belief to be epistemically warranted so as to amount to *knowledge,* on the basis of such practical reasons. As the subsequent discussion will reveal, one of the major advantages of the view I will propose is that it can show us a promising path to answering this challenge.

There are two general schools of thought as to how best to explain these connections. The first position, motivated by the sorts of pressures sketched above, explains the connection by positing a very close relationship between intentions and beliefs about our actions, namely the relationship of identity. This is the "cognitive" view of intention, which holds, at least in its simplest forms, that your intention to φ just is your belief that you will φ.[11] The explanation for why you must believe that you are acting, when acting intentionally, and why the considerations for which you predict are so similar to the reasons for which you intend is that your intention and your belief that you are or will act are one and the same.

Many have remained unpersuaded. This is in large part the result of equally strong pressures in the opposite direction to see intention as distinct from belief. One strong set of pressures to keep intentions and beliefs about our actions distinct stems from the different roles the two types of mental states play in our larger mental architecture. A belief that I will

---

[10] A point made much of in the cognitivist account of David Velleman. See especially Velleman, (1989) Ch. 3.

[11] Perhaps the most ambitious attempt to give such an account is that of Velleman (1989). Others support some weaker version, on which intention is partly constituted by belief, (see e.g. Harman (1979) and Setiya (2008))

φ may affect my inferences and actions in a very different way than an intention to φ.[12] My belief that I will smoke might prompt me to avoid convenience stores and so avoid likely temptation; my intention to smoke, in contrast, will prompt me to go to the convenience store as a means of satisfying the temptation to which I have already given in. My intention to go to the gym might be formed precisely *because* of my belief that will likely succumb to laziness, in hopes of bolstering my resolve.[13]

A second and important set of pressures to keep our intention and belief distinct has to do with the evaluative conditions for believing and intending. My belief that I will act is correct in case it is true; my intention to act is correct in case it is choice-worthy or good.[14] My intention to pick up my friend at the airport might thus count as correct even if the attempted pickup is foiled by an unexpected flat-tire, though my belief that I would pick up my friend was mistaken. Conversely, a skilled pickpocket might, in acting immorally, succeed in believing correctly, but fail in intending well. Insofar as these standards of correctness for one attitude are exclusive, it seems as though the difference in the potential evaluation of my intention and belief provides further pressure to keep the two attitudes distinct.[15]

Thus the second, inferrentialist school of thought, on which the relationship between intentions and beliefs is a much weaker one. On the inferrentialist view, an agent's intention to φ is fully distinct from her belief that she will φ. What accounts for the connection between belief and intention is not some conceptual fact concerning the nature of the two

---

[12] Bratman (1987)

[13] Holton (1999); Holton (2009)

[14] A difference sometimes cached out in terms of differences in direction of fit (see, e.g. Humberstone (1992); Anscombe (1957)) and differences in terms of constitutive aim or function (see e.g. Velleman, (2000); Wedgewood (2002); Hieronymi (2005); Shah (2003); Burge (2003)). I discuss the aim of Intention in more detail in the Appendix.

[15] Though less prominent in the current literature, this second response is in fact the oldest objection to cognitivism, going back to Aristotle NE 3.2

mental states, but an epistemic connection: the intention can provide evidence for, and thus the basis to infer a belief about, the fact that one will φ.[16]

This view satisfies our intuitions about the different roles of an intention and belief that we will act, while going at least some way in explaining the connections between intention and our beliefs which had made cognitivism attractive. We will usually have a belief that we will act when we intend to act, because we will usually be in a position to easily infer from our intending to act that we will act. And we can account for some of the 'non-observational character' of our beliefs about our actions. For while our beliefs are predicted from evidence, like any other, we do have special first-personal access to the intention, on the basis of which we infer the belief that we will act. The apparent immediacy and authority of our first-personal beliefs about our actions can be off-loaded onto the special authority and immediacy of our self-knowledge of our own minds.

But the inferrentialist view captures the epistemic phenomenon imperfectly. First, inferrentialism fails to fully capture the necessary connection between belief and intention. If an agent infers that she will act from her intention to act, there is always the possibility of her failing to draw the inference, however straightforward it might be. But in at least many, if not all, cases of intentional action, it appears as though there is no space for me to fail to believe that I so act. The lack of space suggests that there are not two separate inferences – one practical inference to what I ought to do and another, from my intention to act to what I will do – but one.

Second, inferrentialism fails to face squarely with the special access and authority an agent has in beliefs about her actions. Though I may be in a better position to know my intentions than you are, you could just as easily infer from the fact that I intend to φ to the

---

[16] Grice (1979); Paul (2009).

fact that I will φ, as I can. Nor can the inferrentialist view accept a robust notion of practical knowledge on which an agent believes she will act on the basis of practical grounds. The agent decides what to do on the basis of practical grounds, but comes to believe that she will do it on the basis of having so decided (and on the basis of her reliably doing what she decides to do).

Most implausibly, on the inferrentialist view, distinctive knowledge of one's actions would be impossible for non-reflective agents, like animals and children. Since such agents lack the concepts necessary to represent intentions, they will not be able to infer their actions from the fact that they intend to act. But such agents do seem capable of distinctively first-personal knowledge of what they are doing, just like more sophisticated agents.[17] Though children and animals do not infer that they will act from their intentions, they do not appear to need to learn about their actions observationally, as a third-person observer might.

§III. Motivations for a New View

These objections are likely not conclusive – presumably defenders of Inferrentialism and Cognitivism have things to say in response. I bring them up here for two reasons. First, to motivate the appeal of some third, less problematic solution. But also in order to look at the *structure* of the kinds of problems inferrentialism and cognitivism face, which I think reveals the shape that such a solution might take.

Inferrentialism and Cognitivism result from these competing pressures which appear to push two ways. Identifying intentions with beliefs about our actions will result in a prima facie violation of our intuitions about the different roles the two types of mental states play in guiding and regulating our activity, while keeping them distinct involves a prima facie violation of our intuitions about the close link, within a first personal point of view, between

---

[17] For further discussion, see O'Brien (20007).

intending and coming to believe that you will act. As with many such situations, each view draws its strength primarily from the perceived weaknesses of the alternative.

Notice certain similarities in the form of these pressures for and against identifying belief with intention. The first group of pressures which push toward identifying belief with intention are all what we might call "upstream" pressures – pressures associated with the effects of a subject's prior mental states and processes on some propositional attitude. They stem from the apparent connections between coming to intend to act, and coming to believe that you will act. In contrast, the second group of pressures against cognitivism, and toward inferrentialism, are all "downstream" pressures – pressures associated with the effects of a propositional attitude on the subject's resulting thought and behavior. They are pressures that stem from the different causal, explanatory, and evaluative roles that an intention and belief with the same relevant content might play in our mental architecture.

Notice too that you will resolve both pressures if you could have two distinct mental states that issue from one and the same deliberation. The fact that they issue from the same deliberation would explain why the belief and the intention are formed on the basis of the same considerations. And the fact that a belief and intention issue from the same deliberation would explain why believing that you will act is a mark of intentional action. But the fact that they are two distinct attitudes both preserves our intuitions about the distinction between belief and intention and explains their potentially divergent effects on our downstream mental lives. If it were possible to have two distinct metal states that issue from one and the same piece of reasoning, it would appear to provide a tidy solution.

The problem with this solution, as we saw in chapter 4.1, is that it does not appear possible to have a belief and an intention which issue from the same piece of reasoning. Part of what it is for my attitude to be a *belief* that p, as opposed to some other attitude like an

imagining that p or a desire for p, is that it be the kind of state which issues from (or is partly constituted by) the conclusion to a piece of *theoretical* reasoning about whether p is so.[18] Whereas part of what it is for my attitude to be an *intention* to φ, is that it be the kind of state which issues from (or is partly constituted by) the conclusion to a piece of practical reasoning about what to do.

While not all beliefs and intentions need issue from some explicit deliberation, the fact that they are the kind of attitude which issued from *some* process which functioned to track the truth, or determine what to do, respectively, is part of what explains why they play the distinctive functional role in our mental architecture that they do and why they are answerable to the distinctive standards of correctness to which they are.[19]

A theory on which an intention and belief both issue from the same deliberation will thus face the following dilemma: for both a belief and an intention to issue from the same deliberation, it looks like it would have to be a deliberation aimed at answering either some theoretical question of what is so, or some practical question of what to do. If it is the former, you have not successfully come to an intention; if it is the latter, you have not successfully come to a belief.[20]

This dilemma can be avoided however, if the same token deliberation could be an instance of both practical and theoretical deliberation. For in such a case, the very same deliberation might issue in both an intention to act and a belief that you will act, while

---

[18] I assume here some broadly functional account of mental states. I say partly constituted by their processes that produced them, because they are also functionally individuated by their 'downstream' rational connections to other parts of our mental lives.

[19] For a sampling of the rich discussion of the nature of this answerability, see Anscombe (1957), Moran (2001), and Hieronymi (2009).

[20] A problem familiar from discussion of believing at will, emphasized by Hieronymi (2005).

respecting the conceptual connections between each attitude, and their respective modes of deliberation.

But such a solution merely pushes the problem back. For how could the same token deliberation be both a practical deliberation and a theoretical deliberation? Our prior formulation of the dilemma already contains the germ of an answer. If states are typed by the kind of deliberations from which they issue, and deliberations are typed by the kind of inquiries at which they are directed, then we would need some deliberation, the resolution of which amounts to the settling, at the same time, of both a practical question of what to do and a theoretical question of what I will do.

In the rest of the chapter I will argue that with respect to beliefs about our own actions, we can do just that. When we are reasoning about our own actions, our deliberation can, at the same time, be directed both toward answering a theoretical question of what is so, and directed toward answering a practical question of what to do.[21] Understanding how such a deliberation is possible can help us unravel the puzzle of how beliefs about our actions might be justified on the basis of reasons for acting.

§IV. Decision as Theoretical Inference

How could the same deliberation be directed both toward answering a theoretical question of what is so, and also directed toward answering a practical question of what to do? The answer, I want to suggest, is that for a proposition, p, representing some state of affairs which we take to be under our control – like our present and future activity – we can

---

[21] Here I want to mark the ways in which the proceeding is deeply indebted to the work of Moran (2001)'s on self-knowledge. I see this account as broadly congenial with, and as providing a possible explication of, a transparency account of self-knowledge (Though not, I think, one Moran himself would accept). Just as we can control what we do by deciding to do it, we can control what we believe, by making up our minds about what is so. I begin to discuss some of the possible implications of this account for self-knowledge of our mental states in the next chapter.

answer the theoretical question of whether p is so by deciding that p be so on the basis of optimistic reasons which show p should be so.

In doing so, we are not doing anything remarkable. I want to argue that when theoretical inference is properly understood, we can see that believing that we will act on the basis of reasons for acting is, in all important respects, structurally identical to more familiar cases of believing for reasons.

It should be uncontroversial that when we inquire into whether some state of affairs is so, we often do so by means of employing some mixture of inference rules and premises. But I think the richness and variety of ways we engage in theoretical reasoning is often obscured. Just as it is a mistake to think of practical reasoning as the mere adding up of reasons or desires with a certain valence and magnitude, its also a mistake to think of all theoretical reasoning as the simple tallying of various pieces of evidence, each of which shows the conclusion more or less likely to a stronger or weaker degree.

Theoretical reasoning is not such a monolithic phenomenon. In fact, there is a great variety of kinds of considerations which we might employ in theoretical reasoning, and a great variety of ways in which we might bring them to bear on the question of whether some proposition is so. To give just a few examples:

(i)   Sometimes we answer theoretical questions by means of induction from particular cases. We might infer from green emerald facts to the belief that all emeralds are green.[22]

(ii)   Sometimes we answer theoretical questions by means of abduction from explananda. We might infer from dinosaur track facts to the belief that dinosaurs existed.

(iii)   Sometimes we answer theoretical questions by predicting from causal antecedents. We might infer from storm cloud facts to the belief that it will rain.

---

[22] Here I assume that enumerative induction is a distinct form of inference from abduction, contra Harman (1965). The essential thing for the present discussion, however, is not *which* of these methods are the truly distinct kinds of inference, but that there *are* distinct kinds of inference, operating on distinct kinds of grounds.

(iv)   Sometimes we answer theoretical questions by constitution from grounds. We might infer from the occurrence of certain troop movement facts to the belief that Germany is at war with France.

The list is not exhaustive. There are lots of distinct kinds of inferences, operating on lots of distinct kinds of grounds. Each is a way of answering a theoretical question about whether p is so, by way of pursuing some other question of evidence, explanation, constitution, or causation.

When you answer a theoretical question by means of some particular method of inference, you thereby incur more than one commitment. You are not just committed to the fact that it will rain, for example, you are also committed to the fact that it will rain *because of* the storm clouds. You are not just committed to the truth of Germany being at war with France, you are committed to such-and-such troop movements *amounting to* a case of Germany being at war with France. If you reason to your belief about the existence of dinosaurs by means of abduction, then if you are criticizable for your commitment to the existence of dinosaurs explaining dinosaur footprints, you are criticizable for your commitment to dinosaurs having existed.

When you have control over what happens, I want to suggest that there is a further method for determining whether some state of affairs is so:

(v)   Sometimes we answer theoretical questions by deciding what will happen, on the basis of our practical reasons. You might infer from the fact that she killed your father to the fact that you will get revenge.

When you do this, you will also incur more than one commitment. You are not just committed to its being true that you will get revenge, you are also committed to getting revenge. Insofar as an intention to φ is constituted by a commitment to φ-ing, this also means that you will also have come to an intention to get revenge.

Figuring out that you will act a certain way by deciding that you will act a certain way is just one more method among our varied repertoire for engaging in theoretical deliberation about what is so. Of course, (v) is distinctive from (i)-(iv) in the following respect: the first four are all ways of answering the question whether p by discovering whether p is so. I am treating p as some independent state of affairs, outside my control, where I am to determine whether it obtains by addressing myself to some question which bears on whether p is likely or probable. In contrast, in (v) I am answering the question of whether p by bringing p about.[23]

But I can see no reason why such reasoning ought not count as genuinely theoretical. Each of these methods seem, in the right circumstances, like equally reasonable ways of determining whether some state of affairs is so. A proposition is no less true because I decided to make it so than because I discovered it was so. Moreover, if we really are in control with respect to the state of affairs we are forming a belief about, it will be *good* theoretical reasoning. If we conclude that some state of affairs will obtain, by deciding that state of affairs will obtain, we will be right.

If it is a reasoning which can be directed toward answering a question of what will happen, and it is a reliable means of determining what will happen, I see no reason why this unique causal direction should matter. Deciding to φ is, in all relevant respects, just one more means of answering the question of whether you will φ, along with all the other evidential means we have at our disposal.

What thinking about the case of belief about action reveals is there is an overlooked way to conclude that p is so, by *deciding* that it will be so. In such circumstances, we will have

---

[23] While there are at least some philosophers who are sympathetic to the idea that questions of what we will do are transparent to questions of what to do, fewer are willing to accept that agents in such cases are engaged in theoretical reasoning. Ancombe (1957), for instance, distinguishes between 'speculative knowledge' in contrast to 'practical knowledge' and Moran (2001) between 'deliberative' as opposed to 'theoretical' reasoning.

answered, at once, both a practical question of what to do and a theoretical question of what we will do. We will thereby have concluded a deliberation which was both practical and theoretical, and so have come to form both an intention to act, and a prediction that we will so act, each with differing standard of correctness, and potential divergent effects on our downstream mental lives.

§V. A Genuine Middle Ground?

On the picture I have put forward, when an agent reasons about her future actions, she sometimes does so by engaging in inquiry about whether to φ. In resolving the inquiry, she will have come to two conclusions. She will have come to see the world as being a certain way, and, by making up her mind about what to do, decided to make the world that way. In coming to those conclusions, she will have incurred two commitments, a commitment to the truth of the state of affairs of her acting, and a commitment to act. She will have thus both come to a belief that she will φ and she will have come to an intention, or decision, to φ. She will have come to a belief that she will φ, insofar as she has answered a theoretical question of whether she will φ, and she will have come to an intention to φ insofar as she has settled a practical question of whether to φ.

One may worry that insofar as I have made plausible a picture where we can engage in genuine theoretical reasoning about whether we will φ by deciding that we will φ, I have thereby failed to describe a genuine alternative to the cognitivist and inferrentialist position. One might doubt that, fully worked out, this view can really steer a middle-ground between the cognitivist and inferrentialist view without collapsing into the one or the other.

This worry might go in one of two ways. First, one might suspect that if an agent can answer the question of whether she will act, by deciding whether to act, that the two

questions are somehow equivalent, at least from the perspective of the agent.[24] From the first person perspective, the question 'whether I will φ' and the question 'whether to φ' are indistinguishable, in the way the questions 'are you going to the store?' and 'is Paul going to the store?' might be equivalent, for Paul, though not for Elizabeth.

On such an interpretation, the view that I can answer the question 'whether I will φ' by way of answering the question 'whether to φ' would amount to a species of cognitivism. If the two questions – 'whether I will φ' and 'whether to φ' – are equivalent from the agent's perspective , then the agent's commitment to φ-ing (constitutive of her intention to φ) *just is* her commitment to the truth of the proposition that she will φ (constitutive of her belief that she will φ).

The account would also, like cognitivism, tie the two attitudes together too closely. As we have seen, it seems plausible that in at least some peripheral cases of difficult action, it is possible to intend to act without believing that one will act, or in the case of foreseen weakness of will, believe that one will act without, at present, intending to so act.

On the other hand, one might worry that if what we mean by answering one question by way of answering another is not that the questions are equivalent from the perspective of the agent, it will mean simply that answering one question will put us in a position to answer another question. Sometimes, the fact that I answer a question a certain way – a psychological fact – might put me in a position to answer some other question, even though the considerations for which I answered the first question do not bear on the second question. E.g. the fact that nothing is funny doesn't show me severe, but the fact that I never take anything to be funny might.

---

24 I believe this may be the view of John Schwenkler (2015).

On such an interpretation, the view that I can answer the question 'whether I will φ' by way of answering the question 'whether to φ' would amount to a species of inferrentialism. Though the fact that φ-ing is the thing to do may not show it likely that I will do it, still the fact that I have concluded that φ-ing is the thing to do may. Since I regularly do what I decide to do, my coming to an affirmative answer to the question of *whether to φ* can put me in a position to infer that I will φ.

Focusing on ordinary cases of theoretical reasoning, however, can help us see how a decision-as-theoretical-inference view might differ from either of these interpretations. Consider a couple important points about the case of abduction from explananda.

First, you could have done the abduction properly, though you come to a false conclusion. It may be that the data really is best explained by H, but in this case, the less likely hypothesis turns out to be the right one. So it is not that what it is for H to be true, and what it is for H to best explain the evidence, is identical.

Second, you can *understand* what it is for a hypothesis to be the best explanation of the data, but false. So to say that you answer the question of whether p is so by way of answering the question of whether p best explains the data does not mean that the question of 'does p best explain the data' and 'is p so' are equivalent or indistinguishable for you.

Third, the two questions are separable. You might come to a commitment about whether p is so, *without* coming to a commitment about whether p best explains the data. For there are many different ways to figure out whether the same proposition is so. You might have come to figure out whether p is so by way of some other method, like induction, perception, deduction, prediction, etc.

Conversely, and more importantly, you might come to a commitment about whether p best explains the data without coming to a commitment about the truth of p. You might

think that though p is the best explanation, not-p is true none-the-less (this is possible in part because you are able to understand that these are different questions).

These are all ways in which a view where we answer a question of what we will do by means of answering the question of what to do differs from an interpretation in which the two questions collapse into one another, from the first-person point of view. But cases of answering one question by means of answering another also differ from cases where answering one question merely puts you in a position to answer a second question.

When you answer the question of whether p by means of abduction, you are not inferring from your affirmative answer to the question 'does p best explain the data?' to the fact that p is so. This is a lesson from the Tortoise and Achilles. The application of the rule is not the same as, nor does it require, the representing of that application as a premise in your reasoning.

You can follow a rule without being *able to* represent it. And you can apply a rule well without articulating, and even without having the concepts necessary to articulate, *why* the rule is a good one.[25] I don't need to have the concept of induction, and certainly don't need to know why induction is a reliable method of arriving at the truth, in order to *use* inductive reasoning. In fact, if I infer from the fact that p has inductive support, and the fact that induction is reliable, to the fact that p is so, I am no longer doing induction, but rather some different form of inference altogether.

I return now to the case of beliefs about our actions. The analogy to the case of abduction can help make clear how treating the decision to φ as a means of determining whether you will φ differs from a cognitivist and inferrentialist interpretation.

---

[25] An important lesson from the work of Tyler Burge on Entitlement (see Burge (1993); (1996); (2003)

It differs from a view on which the two questions are reduced to one another in two ways. The most obvious is that not all theoretical reasoning about what is so will be practical reasoning. While a decision on the basis of practical reasons is one possible means by which you answer a question of what is so, it is not the only means. You can answer the question of what you will do by deciding what you will do or by consulting evidence, just as you could answer the question of what is in your pocket via memory or perception. To use Anscombe's way of speaking: they are two ways of knowing, but not two different objects of knowledge.

Equally importantly, not all practical reasoning about what to do will be theoretical. Rather, my view is that practical reasoning, *when* it is employed as a means of figuring out what is so, is a species of theoretical reasoning, along with abductive reasoning, inductive reasoning, etc. But while you can sometimes employ a decision as a way of answering a theoretical question, you need not always employ it so. Sometimes you can just decide what to do, without doing so as a means of answering a question about what is so, just as you might idly reason about whether p best explains the data, without coming to a conclusion about whether p is so. Other times, you might answer the question of what you will do by employing some other set of rules or premises. You might determine what you will do by consulting your evidence, or inferring from your prior brain states, or whatever. You might, at the same time, decide what to do, and your decision what to do comes apart from your conclusion about what you will do.[26]

Like the example of abduction done in isolation from further practical reasoning, these cases of 'pure' practical reasoning will be, admittedly, rare or peripheral. In most central cases, a well-functioning believer will usually be answering the two questions

---

[26] E.g. In Davidson Carbon-Copy Cases. For nice discussion of the problems involved in cases of *akrasia* and difficult action, see Marusic (2012).

together. But that is to be hoped for. In most central cases, we have practical knowledge of what we are doing, or will do. The deliberative view can help explain why this is so.

But determining whether you will φ by deciding to φ also differs from determining whether you will φ, and then treating the fact that you have decided or intended to φ as grounds for predicting that you will φ. First, determining whether you will φ by deciding to φ is a capacity available to all agents, including those without mental concepts. Just as you don't need to represent yourself as performing the mental act of abduction in order to adduce, you do not need to think of yourself as deciding in order to decide – you just need to decide. The deliberative view can thus explain why animals and small children have a distinctive agential way of coming to know what they are doing, as a view in which agents infer their acting from their intending cannot.

Second, when you employ an inferential method like abduction as your means of determining whether some proposition is so, there is no space between the successful inference, and the coming to the belief, as their might be when you treat the fact that p has abductive support as a premise from which you infer (or not) that p is true. So too, though you need not always employ decisions as a means of determining what is so, *when* you do so, there will be a necessary connection between your deciding to φ and your believing that you will φ.

So, I think, the deliberative view *does* manages to succeed in carving out a genuine middle ground between the cognitivist and inferentialist view, without collapsing into one or the other. In doing so, it is able to avoid the problems facing each of these views. The deliberative view can explain, as a theory on which we infer the belief cannot, why a belief that I am acting is sometimes necessary condition for intending that I act. If, in the conclusion of one and the same deliberation, you both come to intend to act and believe that

you will act, there is no room for failing to draw the appropriate epistemic connections between intending and acting. And it can explain, as a theory on which we infer the belief that we will act from an intention to act cannot, our distinctive first-person access. Because I am not in a position to control how another agent acts, I cannot come to believe that she will act by deciding how she should act.

But the view can *also* explain, as a view which identifies belief with intention cannot, why, in certain special cases, like cases of difficult action or weakness of will, we can believe that we will act without an accompanying intention to act, or intend to act while doubting that we shall succeed. For we have the capacity to settle the question of whether a proposition is so by deciding that it is so, only when we take ourselves to have control over what will happen. And in cases of difficult action or weakness of will, we lack such control, and so the questions of what we will do, and what we would like to do can come apart.

§VI. Conclusion

In Part II of this chapter, I have argued that we can understand our distinctive practical knowledge of our actions as involving an ability to engage in theoretical reasoning about what we will do by deciding what to do. I have explained how such a view provides us with a way of understanding how believing for optimistic reasons might be consistent with the requirement that theoretical reasoning aim at the truth.

I claimed that this ability to reason about how we will act could be understood as a species of a more general capacity to believe for optimistic reasons in cases where we take ourselves to have control over the state of affairs we are forming beliefs about: when we have control over the state of affairs we are reasoning about, we can determine whether some state of affairs is so by decide that it be so, on the basis of considerations which show it ought to be so.

I believe that this capacity has broader application. I suspect that our self-knowledge of our mental states might also be fruitfully understood as a species of this more general capacity.[27] Just as we can control what we do by deciding to do it, we have control over our mental states, like belief, in the sense that they are responsive to our reasons for believing. And so we can come to believe that we believe some proposition – e.g. that the Red Sox will win the World Series – by making up our mind and believing it on the basis of the epistemic considerations which show we ought to believe it.

In the next chapter, I will suggest that we understand our ability to form self-fulfilling beliefs on the basis of optimistic reasons as an exercise of this same capacity: when we see that what happens depends on what we believe will happen, we are in the position of deciding what is to happen, and so we can figure out whether it will happen on the basis of the considerations which show it ought to.

If I am right, this capacity to believe for optimistic reasons would provide us with an attractive picture of how to reason about those states of affairs effected by our reasoning. It provides us with the resources to do what our evidence cannot, providing us with sufficient reason to come to a belief about whether we will be a millionaire. But it also provides us with the resources to do what a capacity to believe at will on the basis of pragmatic reasons cannot, to arrive at that belief by coming to a conclusion about the the truth of the belief's content. And it does so while providing an explanation of how we could be *epistemically* justified in believing a proposition because it would be good if it were true.

---

[27] I hope to pursue this possibility further in future work.

Chapter 5: The Epistemic Status of Self-Fulfilling Beliefs

§I. The Case of the Eccentric Millionaire

Let us return to the case of the eccentric millionaire: suppose the millionaire credibly offers us a million dollars to believe that we will be a millionaire. How ought we to reason about whether we will be a millionaire? Should we believe that we will be a millionaire? More generally how, when belief in some proposition is self-fulfilling, ought we engage in theoretical reasoning about, and so come to a belief about, that proposition?

a. Summary Of The Argument So Far

In *Chapter 2*, I argue that we will not be able to arrive at a satisfactory answer if we attempt to discover whether we will be a millionaire by weighing our available evidence. Since the belief that we will be a millionaire is self-fulfilling, whether we will be a millionaire depends on whether we believe that we will. And so to discover whether we will be a millionaire, we must discover whether we believe that we will be a millionaire. But since whether we believe that we will be a millionaire is a matter of whether or not we conclude that we will be a millionaire, which is a matter of whether we believe that we will be, we find ourselves stuck in a deliberative circle. Because the relevant evidence is not independent from our present reasoning, attempts to follow the evidence will be insufficient for our reasoning to successfully arrive at a conclusion which might provide any degree of confidence about whether we will be a millionaire.

Insofar as theoretical rationality requires that we attempt to discover whether a proposition is true by following the evidence, cases where a belief in some proposition is self-fulfilling for an agent will be cases where it is rationally impermissible for that agent to believe, disbelieve, or suspend judgment about that proposition.

Such a circumstance is not impossible. There are, plausibly, cases where the standards of rationality rules out as impermissible any doxastic attitude for certain subjects with respect to certain propositions: (e.g. it is raining but you do not believe it.) But this would be a very odd situation for the standards of rationality to leave us in with respect to self-fulfilling beliefs. For, as I show in *Chapter 1*, there are actual and important cases where our beliefs can affect whether what we believe is true, and we often need to (and appear able to) reason successfully about them.

It would also be an odd situation to find ourselves in because unlike self-defeating cases, self-fulfilling cases are those in which, whatever we believe, we will be correct. If we were to believe that we will be a millionaire, we would be; and if we were to disbelieve that we will be a millionaire, we wouldn't be. It is particularly strange to think that the standards of rationality would forbid us from permissibly adopting any attitude because of a *surplus* of correct attitudes we might arrive at.

But if we cannot permissibly adopt an attitude toward whether or not we will be a millionaire by following the evidence, what are we to do? One promising route, from the work of William James, focuses on a very different class of reasons than our evidential considerations, viz., those considerations which show our being a millionaire to be good or choice-worthy. We somehow break the deliberative circle by employing those considerations which bear on the merits of being a millionaire. Unlike our evidence, which is insufficient to show that we will or won't be a millionaire, there are sufficient reasons to show being a millionaire to be good or worthwhile. And so we would be able to permissibly believe that we will be a millionaire, if reasons sufficient to show being a millionaire to be good or choice-worthy were also reasons sufficient to permissibly believe that we will be a millionaire.

But how could such considerations ever be appropriate to employ in theoretical reasoning, in such a way as to meet the standards of rationality and permit believing that we will be a millionaire? David Velleman suggests an intriguing possibility: that the reasons which show being a millionaire to be good or worthwhile are the grounds by which we decide to believe that we will be a millionaire. Taking up the insight that the result of our reasoning, whatever it is, will issue in a correct belief, Velleman argues that in such special cases rationality allows us to exercise the 'epistemic freedom' to permissibly choose which correct belief to adopt.

For Velleman, when we learn that whatever we believe will be true, it is as though the millionaire had granted us the power to decide what to believe. Because we know that whatever we end up believing will be true, we can see the belief as 'up to us' and so believe at will, on the basis of pragmatic reasons showing such a belief good to have.

Because the belief is self-fulfilling, facts which show the believed state of affairs to be good or choiceworthy will be among these reasons (As I argue in *chapter 4.1*). The fact that with a million dollars I can move to Tahiti to pursue my passions, for example, counts in favor of believing that I will be a millionaire because my believing that I will be a millionaire will make it the case that I become a millionaire.

In *chapter 3*, I argue that as with attempts to follow our evidence, Velleman' strategy will also fail to provide a satisfactory solution to the Millionaire case, because the standards of theoretical rationality will not allow us to believe at will, even in cases where the evidence under-determines a permissible doxastic attitude. Believing at will is not a way of navigating the rational blind-spot we find ourselves in when the evidence under-determines what to believe, but rather a way of circumventing it. To permissibly resolve our theoretical deliberation when our evidence leaves us in a deliberative circle, we would need

considerations which suffice for us to come to some conclusion about whether or not we will be a millionaire. But pragmatic reasons for getting yourself to believe one way or the other – though they might be guaranteed to provide you with a true belief that you will be a millionaire – are not sufficient to show the proposition 'that you will be a millionaire' true.

The lesson from *chapter 3* is that any satisfactory solution to the millionaire case will have to be a solution which shows that the reasons for which we come to believe that we will be a millionaire are reasons by which we settle the theoretical question of whether we will be a millionaire. But if this is the task, how could employing facts about the merits of being a millionaire help you in your reasoning? The fact that you could move to Tahiti if you had a million dollars appears to bear on your decision to believe you will get a million dollars. But it doesn't appear to bear on whether or not it is true that you will be a millionaire.

In *chapter 4*, I argue that we can acquire the philosophical structure we would need to resolve the problem by broadening our conception of what it means for a consideration to bear rational connections to the truth of a proposition.

We might understand such a connection in terms of some e*vidential support* relationship which each consideration (or set of considerations) bears toward a proposition, showing that proposition more or less likely by various degrees. So understood, it is hard to see how a Jamesian solution might ever be implemented – for the fact that you could move to Tahiti if you had a million dollars does not provide such support.

I argue that this picture – where we understand what it is for a consideration to bear a rational connection to the truth of a proposition in terms of evidential support relations – is too simplistic. We should not think of theoretical reasoning as the mere weighing of reasons, each with a certain valence and magnitude, summed to determine how likely or

probable we find a particular proposition. The problem with such a view is that reasons are reasons in virtue of being *employed* in reasoning. Whether, how, and to what degree a set of considerations bears on the truth of a proposition is context-dependent – it depends on the particular details of how those considerations are being employed. In *chapter 4.2*, I suggest we adopt a different picture. Roughly: that considerations bear rational connections to the truth of some proposition, p, by being employed by the agent as grounds from which the agent infers whether or not p is so.

On the suggested picture, whether an agent's reasons actually suffice to show a proposition true (and thus whether or not the agent's belief is permissible) depends on two things. First, it depends on whether or not the kind of inference involved was a good one – whether it is the kind of inference which, when properly employed, provides a good or reliable route to the truth. Induction, for example, is a generally reliable method for getting at the truth. Affirming the consequent is not.

Second, it depends on whether or not the inference was properly employed in the particular case – whether the considerations were sufficient or appropriate grounds for the inference you engage in as you reason about whether p is so. Even though induction, when properly employed, is a good route to the truth, you may lack sufficient reason for concluding a proposition is so if your inductive grounds, in this case, were insufficient for proper inductive inference.

What counts as an appropriate ground will vary inference-type by inference-type. In the case of induction, for example, it may be a certain base of particular instances, in the case of abduction, a certain set of expalanda well-explained by the conclusion. Since the standard for what amounts to a sufficient ground for an inference will be internal to each particular kind of inference, the question of when we have sufficient reason to conclude that p is so

will have no general answer. It too will be a matter of whether or not an agent's reasons were, in each case, sufficient for the inference in which they were employed.

If we adopt this view, we have a new way of thinking about whether a consideration can be a reason for belief: whether the fact that you could move to Tahiti if you had a million dollars bears on the question of whether it is true that you will be a millionaire depends on whether facts which show it good to be a millionaire can be the appropriate grounds for a warrant providing inference to the conclusion that you will be.

In *chapter 4.2*, I argue that they can be. Optimistic reasons can be the proper grounds for a distinctive form of reliable, truth-directed, inference: determining what is so by deciding what is to be so. Deciding what is to be so (when we really do have control over the state of affairs we are reasoning about) is a good route to the truth. And the appropriate grounds for deciding something is to be so are those practical considerations which show that it would be good or choice-worthy if it were so.

Since these considerations are sufficient grounds for deciding that a state of affairs be so, and since decision is a warrant-providing inference, taking those considerations as the grounds for your decision that p is a way of permissibly concluding that p is so, and so permissibly believing that p.

b. The Agent's Reasoning

So how, in the end, are you to think about whether you will be a millionaire, when you learn that you will receive a million dollars if you believe that you will be a millionaire? And how, more broadly, are you to reason about whether p is so when you discover that your belief that p would be self-fulfilling?

Before the millionaire comes to your door (and assuming you have no other route to riches), when wondering about whether you will be a millionaire, you had to follow the

available evidence. You weigh the evidence for your turning out to be a millionaire or not: What are the odds of your winning the lottery? What is the likelihood of your receiving the necessary salary-raises? Do you have any distant wealthy relations? On the basis of this evidence, you might have predicted that it is unlikely that you will turn out to be a millionaire, and so conclude that you will not be.

Upon seeing the state of affairs of your being a millionaire as up to you, you acquire a new way of reasoning about whether you will be a millionaire. Velleman suggested that when we discover that a belief that p is self-fulfilling, we come to see our belief as under our practical control, so we can decide what belief to have on the basis of pragmatic reasons showing the belief good.

I want to suggest, instead, that when a belief that p is self-fulfilling, it is the *object of* belief – the represented state of affairs – which we can come to see as under our practical control. When you are told that you will be a millionaire if you believe that you will (and so learn that your belief that you will be a millionaire is self-fulfilling) you learn that whether you will be a millionaire is up to you.

Once you see it as up to you, rather than predicting whether you will be a millionaire you can now decide whether you will be a millionaire. So instead of asking yourself what the likelihood is that you will turn out to be a millionaire, you can ask yourself: shall I have it be that I be a millionaire?

You weigh the reasons for being a millionaire and for not being a millionaire. If you are a millionaire, you can help out your friends and family, travel the world, donate to charity, take time to write your novel, etc. So you conclude that you will be a millionaire, on the basis of those reasons for becoming one. Your reasoning involved an inferential

transition from the facts that showed you ought to be a millionaire to the conclusion that you would be a millionaire.

This is the end of your reasoning. It issues in a decision to be a millionaire, and, in so deciding, a conclusion that you will be a millionaire. In concluding that you will be a millionaire, you will have thereby come to form a belief that you will be a millionaire. Having formed the belief, your belief will register on the millionaire's mind-reading technology. She will give you a million dollars, making it the case that you become a millionaire. Your belief is true.

Moreover, the reasoning by which you arrived at the belief was good reasoning. You were employing what was, under the circumstances, a reliable inferential-method of figuring out whether you would be a millionaire: deciding whether or not to be. Moreover, you did the inference well. Your considerations really were sufficient for deciding that you would be a millionaire. Since you came to believe that you would be a millionaire on the basis of an adroit employment of a reliable form of inference, your belief is warranted.

Since your belief was warranted, correct, and correct in virtue of the the adroit performance of a reliable type of inference, you seem to be in a good position to count as *knowing* that you will be a millionaire, and knowing it on the basis of those considerations in virtue of which you were justified in wishing to be a millionaire.

c. What Remains to be Shown

In *chapter 4.2*, I argued that we can understand our distinctive practical knowledge of our actions as involving an ability to engage in theoretical reasoning about what we will do by deciding what to do. I claimed that this ability to reason about how we will act could be understood as a species of a more general capacity to believe for optimistic reasons in cases where we take ourselves to have control over the state of affairs we are forming beliefs

about: when we have control over the state of affairs we are reasoning about, we can determine whether some state of affairs is so by decide that it be so, on the basis of considerations which show it ought to be so.

I take myself to have established already that *insofar* as the objects of self-fulfilling beliefs can be understood as under our control, our general capacity to believe for optimistic reasons would help us reason successfully about them. What remains to be shown is that when a belief that p is self-fulfilling, p really can be understood as representing a state of affairs under our practical control.

This step is not without difficulties: self-fulfilling belief differs from the paradigmatic cases of practical knowledge discussed in *chapter 4.2* in an important respect. Our becoming a millionaire is not an action I can perform, and so it is not obvious in what respect we can see our being a millionaire as up to us, and so the proper object for decision on the basis of optimistic reasons.

While I don't have a decisive argument that we must see the objects of self-fulfilling belief as under our control, in the next section I will present a suggestion of how we might think about the scope of our practical control according to which the object of self-fulfilling beliefs might plausibly fall under it, even though it may not involve an action we can perform.

§II. Self-Fulfilling Beliefs and Control

I have argued that we have a capacity to believe for optimistic reasons which extends to anything that falls within my practical 'deliberative field.' That is, for any state of affairs where I take myself to be able to decide whether it is to happen – any state of affairs I take to be under my control – it is theoretically rational for me to believe that it will happen by

deciding that it will. But what is the extent of this deliberative field, and does it encompass the objects of self-fulfilling beliefs?

While the notion of control and the division between those states of affairs over which we have control and those over which we do not is notoriously difficult to specify, there are clear examples of cases on one side or the other. An ordinary intentional activity – like my going for a walk in the park – will, in usual cases, be an example of a state of affairs under my control. My coughing fit, or a state of affairs to which I am entirely disconnected – e.g. that Napoleon's favorite color was blue – will be clear examples of states of affairs not under my control.

There are also less clear cases. My addictive behavior – though intentional – might plausibly be the kind of action with respect to which I am not in control. My beliefs and emotions, though not under our voluntary command, are often thought to be the kinds of things under our sphere of control, as opposed to states, like the coughing fit, which merely befall us.

This broad concept of control I have in mind is closely tied in with concepts like agency, reasons-responsiveness, the active/passive divide, and self-identification. States of affairs we have control over typically involves those states of affairs which we identify as ours; states of affairs we are agents with respect to; those states of affairs for which we are responsible; states of affairs which are responsive to our reasons. While I cannot provide an analysis of the concept here, the division is hopefully sufficiently familiar that we can now ask ourselves where the objects of self-fulfilling belief fall on the divide.

It is not clear that the object of self-fulfilling beliefs should fall under our sphere of control. As we saw in chapter 1, many self-fulfilling beliefs will be self-fulfilling in circuitous ways, often because some third party ensures that the world aligns with our beliefs. In the

millionaire example, for instance, it seems as though it is the *millionaire*, not us, who controls whether or not we will be a millionaire (for all that our belief is self-fulfilling because of how the millionaire will exert that control). When S's belief that p is self-fulfilling, it is true that S will affect whether or not p is so. But such cause-and-effect connections do not show that the causing subject has control over the effect. The fact that some state of affairs is effected by whether S hiccups might show that state of affairs depends on S, but does not show that S has control over it. Similarly, since we lack any kind of obvious direct practical control over our self-fulfilling beliefs, its not clear how a belief's being self-fulfilling gives us any direct control over the believed state of affairs.

When a belief is self-fulfilling we do not appear to enjoy the typical sort of agential control over the state of affairs we are forming a belief about that we do over actions or events which we will or intend. So what is the sense in which, when we see that a belief is self-fulfilling, we see the object of belief as 'up to us?' I want to claim that the relevant sense in which the object of self-fulfilling belief is 'up to us' involves not willing, but wishing.

To argue for this claim, I want to return again to the similarities between wishing and willing, discussed in *chapter 4.1*. Here is Aristotle on the differences between wishing and willing:

> but [decision] is not wish either, though it is apparently close to it. For we do not decide on impossible things – anyone claiming to decide on them would seem a fool; but we do wish for impossible things – for immortality, for instance – as well as possible things. Further, we wish also for results that are not achievable through our own agency – victory for some actor or athlete, for instance. But what we decide on is never anything of that sort, but what we think would come about through our own agency. Again we wish for the end more, but we decide on things that promote the end. We wish, for instance, to be healthy, but we decide to do things that will make us healthy; and we wish to be happy, and say so, but we could not appropriately say we decide to be happy, since in general the things we decide on would seem to be things that are up to us.[1]

Aristotle pulls out three important strands of wishing which distinguish it from willing. First, we can wish for impossible objects, second that we need not reason about means when we

---

[1] Aristotle (1985): NE 3:iii

wish, and, third, that unlike willing, our wishing need not be restricted to states of affairs achievable through our own agency.

These three aspects of wishing are connected. (i) and (ii) will follow from (iii), as impossible objects are, *a fortiori*, not objects which are achievable through our own agency, nor are they the kinds of objects for which there are any means for us to take so as to bring them about. (i) and (ii) are similarly connected. If the object of wishing need not be achievable through our own agency, then we need not have available means for achieving it.

Despite these dissimilarities, Aristotle's wishing is still closely related to willing. The kind of wishing Aristotle picks out is still, in an important sense, practical. It is aimed at action, in the same way that intention is.

One way this volitional aspect of wishing is demonstrated is in the conceptual possibility of akratic wishing: you might believe that p would be best, but wish for some alternative state of affairs, q, instead. You also have a certain practical commitment toward what you wish for, as the thing to happen, in a way you do not, or need not have any practical commitment toward states of affairs you merely believe ought to come about. Consider a case where you wish that the Red Sox win the world series. You are, of course, not an agent with respect to the series. But you are also not just representing the Red Sox as the team that should win the world series. You are 'rooting for' that outcome, and answerable for it.[2] You might even believe that some other team is more deserving, an ought to win instead, yet still, for all that, wish that the Red sox win.

In chapter 4.1, I describe this sense of 'volitional wishing' as roughly equivalent to 'willing off-line.' If willing involves reasoning about what you will have be done, this wishing

---

[2] I take this practical commitment to be stronger than the kind of pro-attitude exhibited by desire in the same way intentions are. One might desire two outcomes equally, but can wish or intend for only one. (c/f Bratman 1984)

involves reasoning about what you *would* have be done. Like willing, this wishing involves more than merely forming beliefs about what states of affairs would be good. But because you are not an agent with respect to what you wish, you do not reason about how to bring it about.

Because you are not an agent with respect to what you wish, your wishing is typically non-efficacious. But it need not always be so. Consider the case of the fairy-godparent who informs you: 'your wish is my command.' Suppose you wish that the Red Sox win the world series. You are still not an agent with respect to the outcome. You need not reason about how to win the world series, or even how to get the red sox to win the world series. But whether the Red Sox win is still, I think intuitively, now up to you.

Thinking about the case of efficacious wishing suggests that there is a spectrum of control, of which the states of affairs we know how to do are only a part. Some states of affairs involve our *basic actions*: actions I know how to do 'just like that' rather than by means of some further action. Other states of affairs involve *mediated actions*: actions I know how to do by way of something else. I can operate the forklift by pressing the button by moving my arm. Our actions can be quite mediated. An architect can build a house by hiring a construction worker who press the button to operate the forklift.[3]

At some point, the connections between your more immediate actions and your aims become so tenuous that the the aim ceases to be an action you are performing, and instead becomes a goal or future action you are working towards. The Architect might be building the house, but the person who orders dinner from the waiter is not making the dinner.

Still, though ordering dinner may have moved out of the agent's sphere of action, it need not have moved out of her sphere of control. The Diner may not be making her meal,

---

[3] Cf Anscombe *(1957).*

but she controls what she will have. And because the dinner is in her control, beliefs about her dinner fall under her capacity to believe for optimistic reasons. Just as the architect can figure out what structure she is going to build on the basis of her practical reasons for building it, The diner can figure out what she is having for dinner by deciding what to have on the basis of her reasons for ordering.

When we know our wishes will be made true, our position is much like the position of the Diner who orders dinner. When the fairy-godparent tells me my wish is her command, I can now decide whether the Red Sox are to win, in the same way the Diner can decide what to have for dinner. And so I can come to know that the Red Sox will win by deciding it be so, on the basis of my optimistic reasons for wishing that the Red Sox win.

We might usefully understand an omnipotent creator's practical knowledge as closer to wishings than actings. God need not know how to make it be light, or create a plan of how to make the light appear. God can instead just decide that there will be light for the reason that it is good, because God's thinking it will make it so.

So too, in cases where our thinking will make it so, I think it plausible that we find ourselves in a position to wish true what we would have so. Consider, from *chapter 1*, the agent whose marriage will be a success in virtue of her believing that it is, or a failure in virtue of her believing that it is. Her situation, I want to suggest, is much like God's. When she learns that whatever she thinks about her marriage can come true, she can now decide whether or not to have a successful marriage.

Though the intuition is strongest in cases where our believing makes its content true in virtue of some constitution-relation, I think there is no reason why the phenomenon should not extend to all cases where our thinking makes it so. Were you to learn that whether the Red Sox win the world series depends on whether you believe that they will, you

are in the position of someone who learns that their wishing the Red Sox win can make it so. Such *effective wishing* – or wishing 'on-line' – is a form of practical inference, with its object under our practical control. Since it is a form of practical inference, we can harness such wishing to determine whether or not the Red Sox will win, by wishing that they do.

Though the objects of our self-fulfilling beliefs are not like actions we can perform or aspects of our mind we can make-up or resolve, I want to suggest that they *can* be well understood in terms of effective wishing. If I am right, then we have discovered a sense in which we can understand the objects of our self-fulfilling belief as under our control, and so the kind of propositions we can permissibly believe because it would be good if they were true.

§III. Conclusion

In this dissertation, I set out to investigate the nature and rationality of self-fulfilling beliefs. I have now concluded my defense of the thesis that, in cases of self-fulfilling belief, we can be justified in believing a proposition on the basis of considerations which show the proposition good. I take this investigation of the rationality of self-fulfilling belief to have taught us important lessons about the nature of theoretical rationality more generally.

First, I hope to have shown that we've failed to appreciate the how deeply puzzling it is to understand what we should believe in cases where our belief determines what happens. I hope also to have shown how this puzzle reveals that we need a more ecumenical notion of theoretical reasoning than we often employ.

Second, I hope to have shown how thinking about self-fulfilling beliefs helps clarify what is most fundamental to theoretical reason which needs to be preserved in any revised theory: that the standards of theoretical reason are grounded in theoretical reasoning's function of determining whether certain propositions are true.

Finally, I hope to have shown how adopting a more ecumenical conception of theoretical reasoning can reconcile the idea that beliefs in certain propositions are justified on the basis of reasons which show that proposition to be good with the idea that theoretical reasoning functions to determine whether the belief's propositional object is true. In so doing, I hope to have provided a new avenue for understanding not just self-fulfilling beliefs, but also claims about the special nature of our practical knowledge of our actions and self-knowledge more generally.

Appendix: The Aim of Intention

It is now widely accepted that belief *aims at the truth* in the sense that beliefs are correct just in case their propositional objects are true.[1] What if, anything, is the analogous aim of intention? In this appendix, I take up the question of how we might formulate an aim of intention and examine two possible candidates: first, that intention aims at the successful execution of the intended action; second, that intention aims at the good. I will show how hewing closely to the aim of belief analogy can help us better understand and adjudicate between these positions, and gain important conceptual clarity about the structure of practical reason along the way. I will argue that a picture according to which intentions are correct just in case the agent successfully executes the action she intended misses important relationships between intentions and an agent's practical reasons for acting. I will suggest that we should instead hold that intention aims, not at successful execution, but at the good. Such a view, I will argue, allows us a more complete way of drawing up the practical normative landscape, while providing a closer analogy between practical and theoretical reason.

What is the analogue of truth for intention? Here is one popular line of response: just as beliefs aim at truth, intentions aim at fulfillment. While beliefs aim at representing the world, intentions aim at shaping the world, producing the action they represent. So an intention is correct just in case the agent successfully executes the act they intended. This distinction between truth and execution is drawn by G. E. M. Anscombe in *Intention*:[2]

> An [intention] will be a description of some future action, addressed to the prospective agent, and cast in a form whose point… is to make the person do what is described. execution conditions for commands correspond to truth conditions for propositions. What are the reasons, other than a dispensable usage for not calling commands true and false according as they are obeyed. (Anscombe 1957: 3)[3]

---

[1] See, e.g. Velleman (2000) or Wedgewood (2002).

[2] Anscombe in this passage speaks of commands generally (whether self-directed or not), but makes clear that she takes the point to apply to intentions.

[3] It is actually somewhat ambiguous, in this passage, whether Anscombe is making a semantic or normative point. But it is clear in later passages (e.g. Anscombe 1957: 56) as well as the subsequent literature on the direction of fit, that a normative point is also intended.

This conception of the aim of intention is grounded in the popular contrast in the 'direction-of-fit' between belief and intention. Whereas the agent is supposed to form beliefs to fit the world, she is supposed to shape the world to fit her intentional representations.[4]

While this account of intention as concerned with the successful execution of action is a natural one, it also seems to leave out a crucial evaluative component of intention. Namely, it leaves out any consideration of whether the executed action was good or bad. Consider the competent pickpocket who intends to steal her neighbor's wallet, and does. Many would like to say that something has gone wrong in her intending, even though she has successfully executed her intended act. But if intention aims at the successful execution of the intended action, what are we to say of our intuition that the wrongness of pickpocketing has an important role to play in normative assessment of the intention to pickpocket?

Anscombe, again, has an answer. While whether an intention is *correct* is a matter of whether the intended action is successfully executed, facts about the goodness or choice-worthiness of the intended action explain whether the intention was justified, or responsible, or rational. Says Anscombe:

> "orders are usually criticized for being sound or unsound, rather than for being fulfilled or unfulfilled; but there is a difference between the types of ground on which we call [intention and belief] sound. The reasons justifying an order are not ones suggesting what is probable, or likely to happen, but, e.g., ones suggesting what it would be good to make happen with a view to an objective, or with a view to a sound objective"[5]

Anscombe here suggests that a view on which successful execution for intention is the analogue of truth for belief can try to accommodate practical reason into the analogy by making use of the distinction, in discussion of belief, between correctness and justification. We can distinguish, with respect to belief, between a belief being true and a belief being justified. Famously, the two can

---

[4] For an overview, see Humberstone (1992).

[5] Anscombe (1957): 4.

come apart: a responsible or reasonable believer won't always believe what is true, because sometimes one's evidence supports a false conclusion.

Consider Matt and Kathy. Suppose Matt believes that dinosaurs continue to exist in isolated portions of Scottish Highlands, despite all his evidence to the contrary. Suppose Matt's belief is, in fact, true: a few dinosaurs survive extinction and continue to live in the freshwater lakes of Scotland. Suppose Kathy, on the basis of her abundant supporting evidence, believes that all the dinosaurs went extinct at the end of the Cretaceous period. Suppose that Kathy's belief is, in fact, false: a few dinosaurs survive extinction and continue to live in the freshwater lakes of Scotland.

Surely something has gone wrong in Matt's believing when he believes that dinosaurs exist despite all the evidence to the contrary, though his belief may nonetheless be true. We can account for this intuition this if we distinguish between *correctness* and justification.[6] In Kathy's case, we can say that the belief, while justified, was incorrect. In Matt's case, that the belief turned out to be correct, though he lacked justification for believing it.

So too, perhaps we can account for the intuition that the pickpocket has gone wrong in her intending, if we can distinguish between the correctness conditions for an intention, which determine when the intention has successfully met its aim, and conditions under which such an intention was justified, or reasonable, or responsible. The pickpocket is like the unjustified dinosaur-believer. Her intention was correct, because her action was successful. But she lacked practical reasons for acting as she did, and so her intending was unjustified, or unreasonable.

A fully articulated view about the aim of intention will thus involve two components. It will specify standard of correctness for the intention, and it will also specify the conditions under which the intention is justified or reasonable. Perhaps, on the fully articulated successful-execution view,

---

[6] For ease of explication, I will avoid issues of whether such justification is external or internal. It is sufficient for our purposes that a belief's justification or lack of justification, whether internal or external, can sometimes diverge from its truth or falsehood.

facts about the goodness or choice-worthiness of an action relates to intention in the same way that evidence for a proposition relates to belief: as reasons for that attitude. As Anscombe puts it, intentions are justified "by reasons for acting, sc. reasons why it would be useful to attractive if the description came true," just as beliefs are justified by reasons which bear on whether their content is true.[7]

But while such a view might have some initial appeal, it faces two crucial difficulties. *First,* it leaves mysterious what the connection is between the correctness conditions of an intention and the justification conditions of the intention. What does an intention's object being successfully executed have to do with practical reasons showing it to be good or worthwhile?

Compare with the case of belief. While there remain deep and interesting issues in articulating the precise connection between truth and justification, that there is some close connection is clear. Roughly: beliefs are justified only if the justification involves some reliable guide to truth. We might hold, analogously, that intentions are justified only if the justification involves some reliable guide to successful execution of the action.

But whether or not an action is practically reasonable is no reliable guide to whether or not the action is successful. Agents often have practical reasons to intend actions which are harder or more difficult than easier, less worthy, alternatives. Conversely, many agents – like the competent pickpocket – may be quite adroit in executing actions that they have strong practical reasons not to do. Moreover, there *does* seem to be a capacity appropriately connected to successful execution of action. But the capacity is not practical reason, but something like the 'cleverness' of the Aristotelian

---

[7] Anscombe (1957): 6.

vicious agent: a capacity involving the technical skill or know-how an agent needs to reliably carry out her goals, whatever those goals may be.[8]

The second difficulty for the fully-articulated successful-execution view involves the need to distinguish between whether an agent had practical reason to perform an action, and whether or not that action was, in fact, the thing for the agent to do. For just as true beliefs and evidence can come apart, so too can facts about which action the agent had practical reason to do, and which action was, in fact, the thing for the agent to do.

Consider the case of Sam and Maggie. Suppose Sam intends to quit his job, leave friends and family, and move to the rural Scottish Highlands, despite all his practical reasons not to. Suppose that moving is, in fact, the thing to do: he would have a wonderful experience in the Scottish Highlands and lead a long, successful, and prosperous life. Suppose Maggie, on the basis of her abundant practical reason not to quit her job, leave friends and family, and move to the Scottish Highlands, intends to stay. In fact, staying is not the thing to do: she would have had a wonderful experience in Scotland and have led a long, successful, and prosperous life.[9]

Notice the strong structural similarities between the case of Sam and Maggie and the earlier case of Matt and Kathy's dinosaur beliefs. Surely something has gone wrong with Sam when he intends to move away despite all his practical reasons not to, even though his resulting action may turn out to be a good one. Just as something has gone wrong when Matt believes against all his evidence in the existence of dinosaurs, even though his belief may turn out true.

Given these similarities, it would be natural, I think, to say that in Maggie's case, her intention, while justified, was incorrect. In Sam's case, the intention turned out to be correct, though

---

[8] See Aristotle NE 6:12 1143b. I thus assume here the broadly Aristotelian picture on which practical reason involves more than 'mere' instrumental reasoning. I do not think this begs the question, however, as it is an assumption which will be widely shared, I suspect, with most proponents of the Anscombian picture I am now arguing against.

[9] I owe these examples to similar cases from Bernard Williams' (1981) discussion of Moral Luck; though I assume in this case, perhaps unlike William's case of Gauguin, it will be uncontroversial that the success of Sam's trip is not of sufficient magnitude to have justified Sam's intention.

he lacked justification for so intending. But notice we cannot say this, if we have reserved correctness of an intention for whether the intended action was successfully executed. For both Sam and Maggie have successfully done what they intended to do. This puts strong pressure on us to put all aspects of normative evaluation in terms of the worth of the action into the *justification* of the intention.

But insofar as we care, in the evaluation of intentions, about whether the intended action was good or bad, we will care along two dimensions. We will care about whether the action was actually good or bad, and we will care about whether, from the perspective of the agent, she had good reasons for taking the action to be good or bad. Placing successful execution of action as the standard of correctness for intention makes us unable to make this important normative distinction.

The fully articulated successful-execution view of the aim of intention thus faces pressure from two directions. If the analogue of truth as the standard of correctness for belief is successful execution as the standard of correctness for intention, then practical reasons seem to be the wrong thing to put in the justification spot. It should be, instead, something about the technical skill, or cleverness, of the agent. Conversely, if practical reason *is* the right thing to put into the justification spot, then successful action is the wrong thing to put in the correctness spot. It should be something about whether the intended action is, in fact, good or worthwhile.

This suggests, I think, that there are in fact two distinct pictures of the aim of intention at work here. One is a picture where intention aims at fulfillment in the sense that an intention is correct if the intended act is successfully execute and is justified by technical considerations showing the action feasible. The other is a picture where intention aims not at fulfillment, but at the good. On this second picture, an intention is correct just in case the intended action was good or worthwhile, and justification will be reserved for whether we had practical reasons for taking the action to be good or worthwhile.

The example of Matt and Kathy helps show the attraction of the second view. The picture where intention aims at the good allows us to maintain the symmetry between the beliefs of Matt and Kathy and intentions of Maggie and Sam. Whether an agent is responsible, or justified, in intending has to do with whether she had good reasons, or justification, for taking her action to be good; it is not determined by whether that action was, in fact, good. Just like whether a belief is responsible, or justified, has to do with whether she has good reasons, or justification, for taking the proposition to be true.

One may worry, however, that this new formulation now suggests the same sort of concerns we had with our initial formulation, in reverse. If intentions aim at the good, what are we to say of our intuition that the successful execution of an action has an important role to play in normative assessment of the intention?

There are, I think, two things we might say in response. First, on a picture where intention aims at the good, there is space for considerations of successful execution in the justification conditions. For practical reasons for acting are not merely considerations which show the action would have good or bad consequences. They also include facts about the feasibility of the intended action. This is most clear in simple 'expected utility' conceptions of practical reason. What action you have most reason to do is a product of both probable success and the utility of the outcome. But it also holds in more sophisticated models of practical reasoning, like those of Bratman or Velleman, on which feasibility of plans will help determine whether, e.g. one ought intend to quit smoking, or intend to take steps to avoid future temptation.[10]

This feature of practical reason will go a long way, I suspect, in capturing our intuitions about the importance of successful action in normative assessment of intentions. Still, there will be cases where an agent has intended to perform an action which was the thing to do and supported by

---

[10] See, e.g. Velleman (1989) or Bratman (1987)

reason, yet fails to do it. Suppose, when push comes to shove, we have to decide between two such options. Compare these two cases:

Good Action: Smith has practical reasons to pick up her friend at the airport, and so intends to pick her friend up at the airport. Picking up her friend at the airport is the thing to do, for the very practical reasons Smith intends. But due to an unforeseeable flat tire, Smith is unable to pick up her friend.

Fulfilled Action: Jones intends to frame her innocent friend for undue profit. Jones is a competent criminal, and succeeds in framing her innocent friend and reaping the undue profit. But framing her innocent friend for undue profit is wrong.

There is, perhaps, something slightly odd about saying, in the first case, that Smith has intended successfully in every respect when her friend remains stranded at the airport. But I think there is something *far more strange* about saying that Jones intended successfully in every respect, while intending and executing bad, perhaps monstrous, activities.

Is there anything to be said in defense of the view that intention aims at fulfillment? I will conclude by considering two possible lines of response we might bring forth in defense of going with Jone's intention over Smith's. The first line of response is to acknowledge that there is a failure in the case of Jones, but to deny that it is a failure of her *intention,* but rather a failure of her as an agent. Jones' *intention* was successful, but *Jones*, in so intending, was a failure.

The problem with such an approach is that there are a wide variety of normative pressures on Jones, and so too many ways in which Jones, as an agent, might succeed or fail. Consider, e.g., cases of so-called 'rational irrationality' like the toxin case, from Gregory Kavka, where an agent is offered a large reward for intending to drink some unpleasant toxin, though she need not follow through.[11] More generally, whether the aim of intention is articulated in terms of the good of the action or its successful execution, we can doctor up a case where the agent does well by intending poorly.

---

[11] For more detailed discussion, see Hieronymi (2005); Kavka (1983); or Parfit (1984)

A second line of response is to wonder why we cannot have it both ways. Why not think there are two different aims or functions – with two different dimensions of success – of intention: one relating to successful execution, one relating to the practical good.[12] The first thing to note is that such a dual-aim view might well be compatible with all that we have said so far, so long as we clearly acknowledge that the practical good of the intended action must be a distinct standard of correctness from the execution of that action.[13] There is nothing obviously incompatible with supposing that intention has two distinct aims, with two distinct correctness- and justification-conditions. This, in just the same way that a belief having some biological or pragmatic function and associated standard of correctness is perfectly consistent with it also having a distinctively *epistemic* standard of correctness.[14]

But though I lack space to do full justice to the issue here, I suspect that a dual-aim view of intention will face serious problems. And that is because once we see the two aims as distinct, we will notice that they can compete in our reasoning about what to do, and so intend. There will be cases where we will be less likely to succeed, precisely because we are choosing the more worthy action. In such circumstances, we will have no way to avoid the dilemma raised by Smith and Jones: one aim will need to have priority. It is my suspicion that it will simply never be the case that intending the easier-to-achieve but practically-worse option is more successful – either more justified or more correct – than intending to do what is practically best.

---

[12] Alternatively, one might wonder why we cannot have some analogue to the knowledge-first accounts of belief. Intention aims at *producing* the good. It can go wrong in two ways, by representing a good and failing to produce it, or by producing something which is a bad.

[13] Or a distinct feature of one complex standard of correctness.

[14] See, e.g. discussion of biological and representational function in Burge (2010): 292 - 308.

Bibliography

Adler, Jonathan E. (1999). "The Ethics of Belief: Off the Wrong Track." *Midwest Studies in Philosophy* 23 (1):267–285.

Anscombe, G.E.M. (1957). *Intention.* Cambridge: Harvard University Press.

Alston, William. "Belief, Acceptance, and Religious Faith" in Howard-Snyder, Daniel & Jordan, Jeff (eds.) (1996). *Faith, Freedom, and Rationality: Philosophy of Religion Today.* Rowman and Littlefield.

Aristotle, (1985). *Nicomachean Ethics.* (*Trans.* Irwin, T.) Indianapolis, Ind: Hackett Pub. Co.

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

Audi, Robert (2008). "Belief, Faith, and Acceptance." *International Journal for Philosophy of Religion* 63 (1/3):87 - 102.

Avveroes (1954) *Tahafut Al-Tahafut*, Van den Bergh (trans.), 2 vols., London: Luzac.

Aquinas, Thomas. (2008). *Summa Theologiae* (2 ed.). NewAdvent.org

Berker, Selim (2013). "Epistemic Teleology and the Separateness of Propositions." *Philosophical Review* 122 (3):337-393.

Bennett, Jonathan (1990). "Why Is Belief Involuntary?" *Analysis* 50 (2):87 - 107.

Boyle, Matthew (2011). " 'Making up Your Mind' and the Activity of Reason." *Philosophers' Imprint* 11 (17).

Boyle, Matthew (2011a). "Transparent Self-Knowledge." *Aristotelian Society Supplementary Volume* 85 (1):223-241.

Bratman, Michael. (1987). *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.

Burge, Tyler (2010). *Origins of Objectivity,* Oxford University Press.

Burge, Tyler (2003). "Perceptual Entitlement." *Philosophy and Phenomenological Research* 67 (3):503-48.

Burge, Tyler (1996). "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society* 96:91-116.

Burge, Tyler (1993). "Content Preservation." *Philosophical Review* 102 (4):457-488.

Burge, Tyler (1988). "Individualism and Self-Knowledge." *Journal of Philosophy* 85 (November):649-63.

Bykvist and Hattiangadi. "Belief, Truth, and Blindspots," In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press. 123.

Byrne, Alex (2011). "Transparency, Belief, Intention." *Aristotelian Society Supplementary Volume* 85 (1):201-221.

Davidson, Donald (1971). "Agency," reprinted in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, pp. 43–61.

Davidson, Donald (1963). "Actions, Reasons, and Causes." *Journal of Philosophy* 60 (23):685-700.

Evans, Gareth (1982). *The Varieties of Reference*. Oxford University Press.

Falvey, K. (2000). "Knowledge in Intention." *Philosophical Studies* 99 (1):21-44.

Feldman, Richard (2006). "Clifford's Principle and James's Options." *Social Epistemology* 20 (1):19 – 33.

Feldman, Richard (2000). "The Ethics of Belief." *Philosophy and Phenomenological Research* 60 (3):667-695.

Feldman, Richard (1988). "Having Evidence." In D. F. Austin (ed.), *Philosophical Analysis*. Kluwer Academic Publishers. 83--104.Philosophical Analysis

Foley, Richard (1991). "Evidence and Reasons for Belief." *Analysis* 51 (2):98 - 102.

Frege, Gottlob & Beaney, Michael (eds.) (1997). *The Frege Reader*. Blackwell Publishers.

Friedman, Jane (2013). "Suspended Judgment." *Philosophical Studies* 162 (2):165-181.

Geach, P. T. (1965). "Assertion" in *Philosophical Review*, 74: 449–465.

Al-Ghazali, (2000). *Tahafut al-falasafia*. M. E. Marmura (ed. and trans.), 2nd. ed., Provo (Utah): Brigham Young University Press.

Goldman, Alvin I. (1980). "The Internalist Conception of Justification." *Midwest Studies in Philosophy 5* 5 (1):27-51.

Gombay, André (1988). "Some Paradoxes of Counter-Privacy." *Philosophy* 63 (244):191 - 210.

Grice, H. P., (1971). 'Intention and Uncertainty,' *Proceedings of the British Academy*, 5: 263–279.

Hall, Richard J. & Johnson, Charles R. (1998). "The Epistemic Duty to Seek More Evidence." *American Philosophical Quarterly* 35 (2):129 - 139.

Hampshire, S. 1959, *Thought and Action*, Notre Dame, IN: University of Notre Dame Press.

Harman, Gilbert., (1976). 'Practical Reasoning,' reprinted in *The Philosophy of Action*, A. Mele (ed.), Oxford: Oxford University Press, 1997, pp. 149–177.

Harman, Gilbert (1973). *Thought*. Princeton University Press.

Harman, Gilbert (1965). 'The Inference to the Best Explanation.' *Philosophical Review* 74 (1):88-95.

Hieronymi, Pamela (2009). "Two Kinds of Agency." In Lucy O'Brien & Matthew Soteriou (eds.), *Mental Action*. Oxford University Press. 138–162.

Hieronymi, Pamela (2009a). "Believing at Will." *Canadian Journal of Philosophy, Supplementary Volume* 35 (sup1):149-187.

Hieronymi, Pamela (2006). "Controlling Attitudes." *Pacific Philosophical Quarterly* 87 (1):45-74.

Hieronymi, Pamela (2005). "The Wrong Kind of Reason." *Journal of Philosophy* 102 (9):437 - 457.

Holton, Richard (2009). *Willing, Wanting, Waiting*, Oxford: Oxford University Press.

Holton, Richard (1999). "Intention and Weakness of Will," *Journal of Philosophy* 96 (5):241-262.

Humberstone, I. L. (1992). "Direction of Fit." *Mind* 101 (401):59-83.

James, William (1895). "Is Life Worth Living?" *International Journal of Ethics* 6 (1):1-24.

Johnston, Mark (1988). "Self-Deception and the Nature of Mind" In Rorty, Amelie & McLaughlin, Brian P. (ed.) *Perspectives on Self-Deception.* University of California Press.

Kavka, Gregory S. (1983). "The Toxin Puzzle." *Analysis* 43 (1):33-36.

Kelly, Thomas (2003). "Epistemic Rationality as Instrumental Rationality: A Critique." *Philosophy and Phenomenological Research* 66 (3):612–640.

Lang, F. R. (2015.) "Control Beliefs Across Adulthood." *The Encyclopedia of Adulthood and Aging.* 1–5.

Langton, Rae (2004). "Intention as Faith." In H. Steward & J. Hyman (eds.), *Royal Institute of Philosophy Supplement.* Cambridge University Press Press 243-258.

Lehrer, Keith & Paxson Jr, Thomas (1969). "Knowledge: Undefeated Justified True Belief." *Journal of Philosophy* 66 (8):225-237

Marusic, B. (2012). "Belief and Difficult Action." *Philosopher's Index* 12 (18):1-30.

McCarey, L., Kalmar, B., Ruby, H., Sheekman, A., Perrin, N., & Marx Brothers. (1933). *Duck soup.* Los Angeles, CA: Paramount Productions.

Moller, "An Argument Against Marriage" *Proceedings of the Royal Institute of Philosophy*, (2003)

Moran, Richard (2012). "Self-Knowledge, 'Transparency,' and the Forms of Activity." In Declan Smithies & Daniel Stoljar (eds.), *Introspection and Consciousness.* Oxford University Press 211.

Moran, Richard A. (2001). *Authority and Estrangement: An Essay on Self-Knowledge.* Princeton University Press.

Morgenbesser, Sidney & Ullmann-Margalit, Edna (1977). "Picking and Choosing." *Social Research* 44 (4):757-785.

Normore, C. G. (1998). "Picking and Choosing: Anselm and Ockham on Choice." *Vivarium* 36 (1):23-39.

Nozick, Robert (1981). *Philosophical Explanations.* Harvard University Press.

O'Brien, Lucy. (2007). *Self-Knowing Agents.* Oxford University Press.

Owens, David (2002). "Epistemic Akrasia." *The Monist* 85 (3):381-397.

Owens, David (2000). *Reason Without Freedom: The Problem of Epistemic Normativity.* Routledge.

Parfit, Derek (1997). "Reasons and Motivation." *Aristotelian Society Supplementary Volume* 71 (1):99–130.

Parfit, Derek (1984). *Reasons and Persons.* Oxford University Press.

Pascal, Blaise, (1670). *Pensées*, translated by W. F. Trotter, London: Dent, 1910.

Paul, Sarah. (2009). "How We Know What We're Doing," *Philosophers' Imprint*, 9: 1–24.

Peels, Rik (2014). "Believing at Will is Possible." *Australasian Journal of Philosophy* 93 (3):1-18.

Piller, Christian (2001). "Normative Practical Reasoning." *Aristotelian Society Supplementary Volume* 75 (1):175 - 216.

Pritchard, Duncan (2005). *Epistemic Luck*. Clarendon Press.

Raz, Joseph (2001). "When we Are Ourselves." In *Engaging Reason*. Oxford University Press.

Reisner, Andrew (2013). "Leaps of Knowledge." In Timothy Chan (ed.), *The Aim of Belief*. OUP. 167-183.

Reisner, Andrew (2009). "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem." *Philosophical Studies* 145 (2):257 - 272.

Rescher, Nicholas (1960). "Choice Without Preference. A Study of the History and of the Logic of the Problem of 'Buridan's Ass'". *Kant-Studien* 51 (1-4):142-175.

Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.

Scanlon, T. M. (2014). *Being Realistic About Reasons*. Oxford University Press.

Scheier, Mf. Matthews, K.A., et. al. "Dispositional Optimism and Recovery from Coronary Artery Bipass Surgery: The Beneficial Effects of Optimism on Physical and Psychological Wellbeing," *Journal of Personality and Social Psychology*, 57, 1024-1040 (1989).

Schwenkler, J. (2015). "Understanding 'Practical Knowledge'." *Philosophers' Imprint* 15 (15).

Shah, Nishi (2013). "Why we Reason the Way we do." *Philosophical Issues* 23 (1):311-325.

Shah, Nishi (2006). "A New Argument for Evidentialism." *Philosophical Quarterly* 56 (225):481–498.

Shah, Nishi. (2003). "How Truth Governs Belief." *Philosophical Review* 112 (4):447-482.

Shah, Nishi & David Velleman, J. (2005). "Doxastic Deliberation." *Philosophical Review* 114 (4):497-534.

Setiya, K. (2011). 'Knowledge of Intention.' In Ford, Hornsby & Stoutland (eds.), *Essays on Anscombe's Intention*. Harvard University Press. 170 - 197.

Setiya, Kieran (2008b). "Practical Knowledge." *Ethics* 118 (3):388-409.

Setiya, Kieran (2008). "Believing at Will," *Midwest Studies in Philosophy* 32 (1):36-52.

Setiya, Kieran (2007). *Reasons Without Rationalism*. Princeton University Press

Setiya, Kieran (2003). "Explaining Action." *Philosophical Review* 112 (3):339-393.

Thompson, Michael. (2011). "Anscombe's Intention and Pactical Knowledge." In Ford, Hornsby & Stoutland (eds.), *Essays on Anscombe's Intention*. Harvard University Press

Thompson, Michael (2008). *Life and Action*, Cambridge, MA: Harvard University Press.

Velleman, David (2000a). "On the Aim of Belief." In *The Possibility of Practical Reason*. Oxford University Press. 244 - 81.

Velleman, David (2000b). "Epistemic Freedom" In *The Possibility of Practical Reason*. Oxford University Press. 32 - 53.

Velleman, David (2000c). "On the Guise of the Good" In *The Possibility of Practical Reason*. Oxford University Press. 32 - 53.

Velleman, David (1989). *Practical Reflection*, Princeton, NJ: Princeton University Press.

Wedgwood, Ralph (2013). "What is the Right Thing to Believe?" In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press. 123.

Wedgwood, Ralph (2002). "The Aim of Belief." *Philosophical Perspectives* 16 (s16):267-97.

Wilkins, William E. "The Concept of a Self-Fulfilling Prophecy," *Sociology of Education* 1976, Vol 49. (April) 175-183.

Williams, Bernard (1981). *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge University

Williams, Bernard. (1973). "Deciding to Believe." In Bernard Williams (ed.), *Problems of the Self*. Cambridge University Press. 136 - 51.

Wilson, G. (2000). "Proximal Practical Foresight." *Philosophical Studies* 99 (1):3-19.

Winters, Barbara (1979). "Believing at Will." *Journal of Philosophy* 76 (5):243-256.