On What We Should Believe

A Thesis Consisting of Interrelated Papers

Subject of Thesis: Epistemic Ethics and Axiology

Nathaniel R. Helms

Oriel College

DPhil in Philosophy

[Word Count (excluding bibliography): 77,177]

Abstract: 5
Introductory Remarks: 6
Chapter 1: Self-Undermining Doxastic States 10
Chapter 2: New Defenses of Volitionism 37
Chapter 3: Pre-Modern Conceptions of Epistemic Value 81
Chapter 4: The Present Context 114
Chapter 5: Breaking New Ground 138
Chapter 6: Strengthening Subjectivism 163
Concluding Remarks 198
Bibliography: 200

ABSTRACT

My aim is to understand our epistemic obligations. To what extent are we able to believe at will? If we are not able to believe at will, what ought we go about investigating? What must we learn for moral reasons? Independently of morality, what is of epistemic value? The dissertation seeks to help build a grand vision of the relationship between ethics and epistemology. It does not propose such a vision; rather, its organizing principle is to fill gaps in the literature to help enable such a vision. The thesis focuses almost exclusively on untrammelled ground. For this reason, some chapters are disconnected from others. However, each chapter presents opportunities and dangers for philosophers working on more widely discussed questions, for each has implications for their subject matter.

For Dad

Introductory Remarks:

This short introduction is a narrative account of what unites this dissertation. Its parts are integrated to a significant degree, but a story about how one part leads to another may help fully to clarify. My project was to form a broad view integrating the ethics of belief with epistemic axiology and to fill gaps in the literature along the way. Because those gaps turned out to be surprisingly large, this dissertation consists almost exclusively of six theses aimed at filling those gaps. My aim in this introduction is to give a sense for how these theses relate to each other. Some connections are tighter than others, a fact I hope is forgivable in view of my goal of advancing the conversation surrounding the ethics of belief in new ways. My priority has been to break new ground rather than to tightly interweave subject matter. With that in mind, I here offer a vision of the motivation for the project. For the purposes of this introduction, claims for which I have argued at length have been put into italics; the rest functions as a frame for the reader.

I began the project by investigating doxastic voluntarism, the view that we can believe at will. I concluded that perhaps with a very strange exception, we cannot. That exception concerns cases in which an agent's believing *p* actually makes *p* true. Since William James there has been some debate about such cases. While they involve forming a belief that is not yet true at the time of deliberation, they terminate in true belief, which is a fairly significant mark of their rational acceptableness. Such cases, I argue, can be inverted in such a way that *no* traditionally acknowledged doxastic stance can easily accommodate, including agnosticism. It is not clear to me what these cases imply exactly, only that whatever they imply is odd. I argue for the disjunction that *either there is at least one doxastic stance beyond belief, disbelief, and agnosticism, or that there are cases in which every doxastic stance is inappropriate*.

Having found little to add to the debate about doxastic voluntarism, save this short chapter explaining the above strange exception, I turned next to the implications of doxastic involuntarism. These were troubling. If I cannot choose what I believe, then either I am not morally responsible for

what I believe, or I am responsible only in virtue of upstream choices, or I am responsible in virtue of something else entirely. A trilemma follows:

1. If we are blameless for what we believe, and our actions are at least very frequently (if not always) the product of our beliefs about what we should do, then it is difficult to see how we can be morally responsible for anything at all.

2. According to "volitionism," choice (or something similar) must underlie all fundamental moral responsibility. On this view, we can suppose we are responsible for our beliefs only insofar as they have been shaped by our previous choices. Well, which choices? And how responsible? Presumably, the answer to the first question is "Only those choices that were themselves voluntary enough to ground moral responsibility."[1] The answer to the second will involve some combination of reasonable foresight and control, for it could not be the case that every choice with downstream epistemic effects is sufficient grounds for praise and blame. If that were the case, happening to idly wander down the wrong aisle of the library could be damning. Importantly, even with sufficient foresight and control conditions, it will be very difficult to tell whether individual bad actions or beliefs are blameworthy.

3. Perhaps our beliefs, while involuntary, disclose our inner selves in such a way as to be already the sort of thing for which we are responsible. Or, perhaps: We are culpable for our bad beliefs not because we have willed them, but because we have thought them. Or, perhaps: The concept of the voluntary has been widely misunderstood, and epistemic negligence or recklessness is no less culpable than other kinds of negligence or recklessness. This family of views simply rejects the basic intuition that we are only directly responsible for what we choose.

To my mind, the second horn is the least bad. It comes at the cost of considerable scepticism about individual cases of responsibility ascription, and that cost is not small. We cannot know whether or to what extent an agent's bad beliefs (and downstream bad actions) are blameless, because we do not know to what extent her beliefs were influenced by free actions rather than elements of her upbringing or genetics or environment. In my second chapter I examine what I have found to be the

only two unanswered strong arguments attacking the "volitionist" horn and offer some considerations in its defense. First, I argue that *Pamela Hieronymi offers an account of agency that is neither persuasive to volitionists nor intuitively viable in its own right.* Second, I argue that *John Hyman's implicit critique of volitionism identifies nothing problematic about volitionism, and, if it does, itself succumbs to* tu quoque *problems*.

The portion of the thesis these two chapters compose does not attempt anything revolutionary in either the ethics of belief or the philosophy of moral responsibility. Rather, its aim is shoring up the defenses of two already popular positions, namely doxastic involuntarism and volitionism. If these positions are correct, we might wonder what conclusions to draw. Part of the answer is clearly that we have to take our epistemic obligations very seriously indeed. If our bad actions are often downstream of bad beliefs, and our bad beliefs are themselves downstream of epistemically-consequential actions, we ought to be very careful to get our beliefs right.

We do this by working on our intellectual virtues. We read widely and we guard against various kinds of irrationality—from wishful thinking, from epistemic neuroticism, from every bias we can identify. We also make sure we spend our limited cognitive resources on the right questions. These will often be moral questions, since the moral life, one assumes, takes priority over the rest.

But what of the so-called deep truths about the nature of reality? Do they play a necessary role in our flourishing? Do they have some kind of moral status? Do we owe them some of our scant attention? Questions like these are not easily answered for all sorts of reasons. One reason, however, took me by surprise. In the process of investigating them, one might, as I did, examine which kinds of things are most valuable to discover, from an epistemic point of view. If so, one will be surprised by the near total absence of contemporary analytic literature on the matter. There is a great deal done in contiguous areas. For example, there is much written about the axiology of various epistemic *states* ("SA"); namely, work discerning which epistemic states are of fundamental value, or of any kind of value. Perhaps all that matters fundamentally (again from an epistemic point of view) is true belief—or perhaps knowledge or understanding. Maybe all are of value in an irreducible way. The subject is widely discussed, and with some impressive results. If, however, we look at questions

concerning the value of different kinds of *content*, we find almost nothing except throwaway lines and gestures about how it must surely be the case that e.g. knowledge of whether there is a God or of the fundamental laws of physics is more valuable than knowledge about patches of sand in the Sahara.

Sorting out epistemic content axiology ("CA") is important because only then can we see what the epistemic realm demands of the ethical realm. I have argued already that ethics makes strong epistemic demands, i.e. that we must try very hard to get the right moral beliefs. But until we can see what the realm of cognitive value demands we will not have an adequate view of their relation. While there are very few views of CA on the market today, there used to be more, and they help orient the rest of the thesis. I will say more about that in a moment, but as an illustration of the connection I am attempting to draw, we can imagine three views:

1. That which is hardest to fathom, but can be known with the deepest understanding, is the most valuable thing to know. Moreover, plumbing those depths is of the highest importance.
2. Much that is important to know cannot be fully comprehended, but remains important to know because it is the heart of reality, and getting to the heart of reality is partially constitutive of human flourishing.
3. We are curious creatures, and the satisfaction of our curiosity, which may or may not get to the heart of anything, is a good thing in and of itself. That which is important to know is that which we want to know.

The views are, roughly and controversially, Plato's, Aquinas', and Hume's, or so I will argue. Each suggests a different approach to the question just above. If Hume is correct, then it is difficult to see why we should take distinctively epistemic axiology, insofar as such a thing exists, very seriously. What claim could curiosity have over morality? No doubt it can maintain a status similar to hunger, insofar as its satisfaction enables other moral endeavors, and can even speed them along, given the right circumstances. On the other hand, the Platonic view legitimizes not just taking epistemic content axiology seriously, but coupling it closely with debates about the axiology of epistemic *states*. The Thomistic view, like the Platonic one, prioritizes getting to the heart of reality and puts such an

activity on a very high pedestal, but uncouples content from states. The most important things to know, on Aquinas' view, diverge from what can be fully understood.

Now, I do not hold that these views logically entail in an obvious way views about what it is morally important to learn. It is not obviously inconsistent to be a Humean who thinks garnering epistemic value to be of great moral importance, for instance. The view, however, is not a natural one. Far more likely, different commitments about epistemic value will have different natural allies in ethics. Coming to grips with epistemic value is, then, part of coming to grips with what we are morally obliged to learn.

After the historical groundwork has been laid out, the remainder of the thesis takes the first steps towards outlining various approaches to the problem of CA, with no special attention paid to historical accuracy or interpretive minutiae. The fourth chapter is in a vaguely Platonic spirit. *I there look into the current literature on the intersection of epistemic content axiology and epistemic state axiology and conclude that very little has been shown*. It is an unfortunate chapter in its negative character, but may spur further research. The good news, insofar as there is any, is that there are no credible indicators that a focus on CA will saddle us with any outlandish views about SA.

The fifth chapter is the very beginning of a defense of the Thomistic view, which is the approach to which I am intuitively attracted. I say the very beginning because I spend most of the chapter explaining how varying metaphysical and even psychological traits pertain to questions about what is at the "heart" of reality. In fact, much of the chapter is spent explaining what it would mean for reality to have a heart, and who should care if it does. This is by far the most experimental section, unmoored from the literature, but I think it attempts something necessary if we hope, as a discipline, to come to grips with what it is for a question to be deep, and in turn to grapple with whether that matters.

Throughout, I am not very kind to the Humean view, or indeed any "subjective" view according to which epistemic value is indexed purely to facts about us, and not about the world. The last chapter attempts to be more charitable. I do not think naïve Humeanism is viable, but I think it

can play a role in the very reductive view, hinted at just above, according to which epistemic value is not enthroned next to the moral (or practical) value of knowledge that we need in our operational lives, but sits below it, as a handmaiden. I do not agree with this hybrid view, but I find it difficult to dismiss totally. My final chapter is an attempt to strengthen the view.

Chapter 1: Self-Undermining Doxastic States:

1.1: Introduction:

In this chapter I examine a particular set of self-defeating doxastic states and put forward a menu of views on what theoretical rationality requires with respect to them. While there is relatively little scholarly literature on this question, there is some on a closely related one first advanced by William James (1896) and developed canonically by David Velleman (1989). Velleman considers cases in which a proposition becomes true only once it is believed, and argues that in such cases it is rationally permissible to believe the proposition in question. Other philosophers argue that we are barred from giving assent to such propositions, because, roughly, they are neither true nor supported by the evidence prior to assent. My focus will be on the opposite kind of case; namely, the kind of case in which the relevant proposition will be false if and because it is believed—and true if disbelieved. Indeed, I go further: If I am right, there are cases in which even agnosticism is problematic.

The chapter is structured as follows. First, I will lay out a central case, and raise some worries about its paradoxical implications. Next, after revising the case to ensure its coherence, I will consider a number of approaches for how to deal with it. In the end, there are two popular theses this chapter's argument threatens. I hope to show that at least one of the following is false:

> Optimism: For any set of evidence E and for any proposition Q, E will, all things considered, support either believing Q, disbelieving Q, or being agnostic on Q.[1]
>
> Triad: There are only three doxastic attitudes—belief, disbelief, and agnosticism—and once you've sufficiently considered a proposition, there are, intellectually speaking, only three options open to you: you either believe it, disbelieve it, or withhold judgment.[2]

---

[1] These two principles are borrowed with minimal changes from Turri (2012, 356-7). Regarding Optimism, see also Conee and Feldman (2004: ch. 4) and White (2005: 445-59).

[2] See also Sosa (1991), Chisholm (1989), Feldman (2002), Pace (2010), Ryan (2010) for a few examples of endorsements (sometimes implicit) of Triad.

It is unremarkable that there are cases in which neither belief nor disbelief are appropriate. It is much more remarkable that there are cases which even can be considered indefinitely, even with all possible evidence furnished, that cannot rationally terminate in belief, disbelief *or* agnosticism.

1.2: Self-Fulfilling Velleman-style Cases:

In his *Practical* Reflection, Velleman argues that it is sometimes rationally permissible to believe that which is made true by one's believing. The purpose of reason, *inter alia*, is to furnish us with true beliefs, and in the sort of cases described, one is obtaining a true belief. For some, our beliefs ought to be responsive to the evidence already sitting before us about what is the case; but on this view, we are less constrained, for our general policy of taking evidence seriously exists only to guide us towards true beliefs. If the truth of the proposition at hand is itself furnished by our believing it, we can take more liberties.[3]

While Velleman developed the canonical formulation of the view, some form of it goes back at least to William James' defense of pragmatism:

> There are, then, cases where a fact cannot come at all unless a preliminary faith exists in its coming. And where faith in a fact can help create the fact, that would be an insane logic which should say that faith running ahead of scientific evidence is the 'lowest kind of immorality' into which a thinking being can fall. Yet such is the logic by which our scientific absolutists pretend to regulate our lives! (1896: 25).

James' great opponent, Clifford (1877), held the view, sometimes called evidentialism, that one ought to proportion one's belief to the evidence. On this picture, I would not be (rationally) permitted to believe that which becomes true only upon my believing. If a runner can win a race if and only if she believes she will win the race (the confidence that her belief constitutes makes her faster, say), then on James' and Velleman's view she is perfectly within her epistemic rights to do so; on the other view, she is not permitted to do so. For, by stipulation, it is not now the case that anything guarantees that she will win, and it is not now the case that she has any evidence that she will win. So, if she were to

[3] Or perhaps: "If sufficient *evidence* of the truth of the proposition at hand is furnished by our believing it." More on this later.

form this optimistic belief, she would not be doing so on the basis of any evidence "now" before her, despite it being the case that the belief would be supported once formed.

This chapter is not directly concerned with Velleman-style cases, though it is indirectly so. Rather, I will examine the opposite kind of cases, those in which whatever doxastic attitude an agent may take will be undermined. It is worth noting that these are not the only two types of cases which bear the sort of self-referential structure under scrutiny. In Velleman-style cases, believing brings about true belief, disbelieving brings about appropriate disbelief, and agnosticism leaves the state of the proposition at hand untouched and subjectively indeterminate—three seemingly acceptable results. In the case that follows, believing brings about the falsity of whatever is believed, disbelieving brings about the truth of the relevant proposition, and agnosticism is rendered inappropriate (or so I will argue) once *that* attitude is embraced. Thus, we have three fairly good outcomes in the first instance, and three bad ones in the second. But these are not the only kinds of cases possible. It might be, for example, that there are cases in which belief comes out poorly, disbelief ends up ok, and agnosticism leaves things untouched. There are plenty of permutations, only one of which we will discuss in detail.

1.3: Self-Undermining Doxastic Situations ("SUDS"):

For the sake of simplicity, I pursue a case-model that is simply a modified version of the Velleman-style case; one that is a reversal of one offered by Peels (2014b). The case is very much of the sort philosophers contrive, but as I will argue later on, we do not completely want for analogues in real life.

One can imagine a timer displaying 60 seconds and a counter displaying "0" sit before a subject. At the start of the experiment, a reading of the subject's doxastic attitude is taken by a trustworthy lab technician with a perfectly accurate brain-scanner. If the subject believes she will win $1000 when the timer goes off, the counter will fall by 1 (in this case, to -1). If she has a different attitude (one of disbelief or agnosticism), the counter will rise by 1. If she has no attitude, the elapsed second is added back to the clock, and nothing will happen to the counter until another reading of her

mental position is taken and a stance is discovered. Readings will be taken every second, and the results added to the tallies. If and only if, at the end of the experiment, the counter displays a positive integer will the subject win the money.

Here is an intuitive, naïve account of how the subject would fare: We might suppose the subject will consider her position before she forms any kind of attitude. This is not to say she is agnostic, for we are not normally agnostic about that which we have yet to consider.[4] Eventually, once she grasps the nature of the game, she may form an agnostic attitude, and the clock will tick down to 59 seconds, and the counter will tick up to 1. For as long as she stays agnostic, she will continue to garner points. But it is unlikely she will be able to run out the clock this way. For, as the count rises, so does her evidence that she is on the road to victory. Once that happens, she will start to become confident, and begin to lose points. This is not to say it would go this way for everyone. For a normal human being, I think there are many ways the experiment could go. Some people might oscillate back and forth around 0. "Look," they would say to themselves, "the counter is rising—I'm on my way to winning!" And, a moment later, "I'm going to win!" At this point, of course, the counter would start ticking back down. Others might be initially optimistic and then start to lose confidence as the counter dips deep into the negative integers. However, most anyone who understands the setup will be epistemically tempted to oscillate.

Different lines of consideration will have different effects on the subject. The subject may see how difficult it is to run up a high point tally (or a low one, for that matter), and reason that agnosticism is warranted. This will help bring in more points (for that is according to the rules we have stipulated). But, just as the subject will notice that rapid oscillations warrant temporary agnosticism about the final result, so she will notice that sustained agnosticism is a path to victory—and thus will be unable to sustain it! She may think she has no way of keeping the tally up, and keep it high, but only until she notices what she is doing.

[4] Or so I assume. More on this later on.

Because of the complex nature of the case and our ignorance of the manifold contingencies of human psychology, it is difficult to predict what real-world agents would do. However, as epistemologists we are also interested in what a rational agent would do, not just an average one baffled by a thought experiment. Nonetheless, getting a sense of the problem is not an easy exercise. It helps to stipulate that she takes this test every morning. Sometimes she wins, we might suppose, and sometimes she doesn't. So, while the slightly confusing set-up might have yielded lucrative agnosticism at first, any strategy she might pursue, as long as she recognizes she is pursuing it, will be self-defeating. A winning program, once recognized as such, will cease to get results.

The point here is that higher-order, level-splitting approaches don't seem to help a rational agent find a stable, rationally defensible strategy. Plenty have been suggested to me. For example, it might seem that a rationally unobjectionable winning strategy would involve sealing off considerations about the evidence for or against her winning a particular point and her focusing instead on inductive considerations. "Look, there's really no way to tell if I'll win this particular round," she would say to herself, "for I haven't had any luck at consistently winning." Her approach is roughly inductive, and we might think we are sometimes licensed to do such things.

In the example just above, she reasons inductively from her past success to the likelihood of her future one. But induction, at least as intuitively understood, is a false friend here. Notice, again, this is true even of the circumspect sort of induction born of repeatedly playing out the scenario. If one maintains that one really has no idea what the outcome will be, one has a seeming recipe for success—but it is one that is unlikely to last very long. It would be a very strange agent indeed who could win, say, thirty scenarios in a row on the basis that she has inductive grounds for a studied agnosticism. The error the level-splitter makes is in thinking that induction here is evidence about her evidence. It is not merely that. It is also evidence about what will happen, and thus the level-splitting approach never gets off the ground. The upshot, then, is not that she can believe whatever she likes, but the contrary—that she cannot form any sort of stance at all.

This, at any rate, is what I argue: that the agent is rationally obliged not to take any doxastic stance (not even an agnostic one). One way to get a sense of why is to consider what might be called

the "tipping point." Each scenario contains a tipping point, at which some doxastic stance or set of stances would guarantee an outcome. With 30 second remaining, a counter reading "30" will be one such tipping point. If the subject can get just one more point, she can relax, for even if she subsequently forms the belief that she will win, not enough time will remain for the counter to drop down to zero. Or, if she has one second remaining, a counter reading "1" will also qualify.

A simpler formulation cuts out the clutter built up from other philosophers' original cases. To win, all the subject has to do is either believe she will lose, or form an agnostic stance. To lose, all the subject has to do is to believe that she will win. So, her options are either to believe something false, or to be agnostic about a proposition for which she seemingly has a great deal of evidence.

This mismatch between anything she might do and the evidence is one problem. Another is that when the subject asks herself, "Will I win?" she may have no decisive evidence to which she can advert, because perhaps she may have no attitude at all. It is true that she knows a number of counterfactuals, but these provide no decisive basis. If we suppose that she has considered these matters for a thousand years, and knows all there is to know about them, and we put her in a stripped-down, one-second-long version of the case, with an immediate tipping point, what's she to do? If the timer only moves once she has taken a doxastic stance, and if she is committed to rationality, will it ever tick down? I will return to these questions in due course.

1.4: Previous work on the problem:

Not much literature exists on SUDS-style cases. Reaber (2012), the only paper I could find on this kind of case, argues that cases of this sort, constituting genuine theoretical dilemmas, might only arise "for hyperidealized agents unlike ourselves," but that cases in the near vicinity can arise for us. As I mentioned above, they need not be particularly contrived. One of his examples concerns Laila, whom you'd like to ask on a date, but who has a distaste for cockiness. Since you know Laila to be perceptive and your own demeanor to be of the sort that betrays your attitudes easily, you find yourself in a bind like the one above. If you believe that Laila will say yes, she won't; otherwise, she will. We might extend the case such that Laila is charmed by agnosticism. If you know this about

Laila, you are in exactly the kind of case described: Whatever stance you form will lead to her going on a date with you—except believing she will go on a date with you.

Reaber goes to great lengths to lay out a probabilistic framework in which to discuss the problem. Because he has already done this groundwork I won't concern myself with it further. He also considers, in some detail, objections to the possibility of SUDS concerning the vagueness of belief, the imprecision of belief, and others, concluding that they fail. Since his responses are convincing, these too I will lay aside. What I will discuss is his argument that SUDS cannot arise for agents like us.

Reaber's discussion centers around a condition C. Here is his description, in full. C has the following characteristics:

> 1) Credentially luminous: if C obtains, your credence that it does is high, and if it doesn't, your credence that it does is low.
>
> 2) A-chance contraindicating at t: if C obtains at t, the chance of A is low, and if doesn't, the chance of A is high.
>
> 3) A-credence loosely covariant: if your credence in A is high, C obtains, and if your credence in A is low, it doesn't, but if your credence in A is middling, C may or may not obtain (816).

The language here is technical and the issues obscure. I will illustrate what Reaber is getting at momentarily. But what Reaber is describing seems to be the same phenomenon instanced above. If C obtains, the agent will believe that it obtains, and if it doesn't, the agent will believe the opposite. This is a problem, for if C obtains at t, the agent will believe that A is unlikely to be true—and if the agent believes A is unlikely to be true, then C will not obtain. A similar paradox will occur for both disbelief and agnosticism.

1.5: Timing Problems:

There seem to be at least two timing problems with SUDS cases. The first Reaber himself addresses:

> Perhaps most interesting is whether the condition that you are sure at t that it is t can be weakened. At least if you are sure that time is continuous, it might be reasonable for you to

> always be sure that it is not t, in which case this whole approach to generating a realistic theoretical dilemma looks happily doomed (816).

Indeed, if you can never be sure it is t, it will be for the same kind of reason that would support the contention that you can never even know that it is t. The idea is something like this: The brain scanning occurs every second, but you do not know exactly when the scan will be taken, and since you need to figure out what you think (or, perhaps, what you will think) when the scan takes place in order to know whether you will gain or lose points, you can default to a comfortable agnosticism without immediately falling into the kind of paradoxical position we have been discussing.

I am unsure what to make of this objection. It may be that it fails an iteration test. Suppose that you think that you can never know whether it is t. If this is the case, by stipulation we can make it such that you will win money whenever you have an agnostic attitude about whether you will win money. Initially you will do very well, but it would be remarkable if you could win an iterated game like this one every time. At some point you will fall back into instability, and some version of the paradox will reemerge. But even if I am wrong about this, I suspect this objection can be sidestepped with a small, tidy change. Let us suppose that the subject's brain is scanned constantly, and that the scenario terminates straightaway once she has taken any stance. That is to say, she will win whatever sum immediately, or lose immediately, without further iterations. It now looks like she doesn't need to worry about when the scans will occur, and what implications, if any, that timing question has.

But the other problem is much more interesting. Reaber is supposing that several things are true at once. Let's take example of how Reaber thinks things could go. C and A are at odds in that C indicates the falsity of A, and C is conditional on your believing A. If you believe that C obtains, it doesn't, so you believe something false if you do. Since the other options have the same problem, you seem doomed to irrationality. The problem, in a nutshell, is that it is not possible for you to know that C obtains at the same time that you infer A, for inferring A makes it the case that C, the basis of your very inference, is false.

An example helps. Suppose you are in a timed scenario as described earlier. Here, you win a sum of money just in case you do *not* believe that you will win the money when a timer goes off. Any

other stance is a losing one. Further suppose that there is a visible sign that reads "WINNING" in large red letters just in case you lack the belief that you will win (for it is only if you lack this belief, when the timer goes off, that you win). If it is lit when the timer goes off, you win the money. You know all of this. On Reaber's picture, and on the one I have sketched, this sign exposes a kind of paradox. For the sign is a reliable indicator that you are winning, and seemingly provides good evidence that you are winning. After all, by stipulation, it always tells the truth. For agents committed to rationality, it seems like incontrovertible evidence that you are winning, and for this reason it ought to be believed. On this picture, all of the following are purportedly true at the same time, the time you make an inference and thereby first come to have a belief:

1) The sign, which you know is infallible, is lit.

2) Whenever the sign is lit, you know the sign is lit.

3) You know that the lit sign indicates that you are winning.

4) You competently infer, from the sign or from your knowledge of the case, that you are winning.

For simplicity's sake I will focus on the case where the sign is lit only, but the following inference has all the same features:

5) The sign, which you know is infallible, is dark.

6) Whenever the sign is dark, you know the sign is dark.

7) You know that a dark sign indicates that you are not winning.

8) You competently infer, from the sign or from your knowledge of the case, that you are not winning.

Reaber seems to think the inference (e.g. that you are winning) is obligatory, and *a fortiori* permissible. It certainly seems that it should be, for we make such inferences all the time. The sign is lit, a lit sign means I'm winning, so I must be winning. This is a simple application of Modus Ponens.

From there, the scenario would play out as it did before. Upon making the inference that one is winning, one ceases to be winning, and perhaps begins to think that one will lose. Then, as is by now familiar, the sign goes back on. As time goes on, one may come to be agnostic about which state one will be in when the timer goes off. But, as we have seen, such agnosticism is itself unstable (for we know that it leads to victory). What, then, is one supposed to think? As time slips by, it remains very unclear what attitude is rationally permitted.

But Reaber's analysis goes wrong at the start. Given the strangeness of the story just above, the reader may already have some reason to think that something is amiss. The clearest way to see what is wrong is to consider what would happen to the sign at the time of an inference. Now, recall that the sign is merely a visual indicator of what you already have access to. When you are in C, by stipulation, you know you are in C. At precisely the time the inference occurs at (4); that is, when the agent says to herself, "The sign is lit, so I'm winning," either the sign is on or it is off. Suppose it is off. Then if follows that:

9) When the inference is made, it is not made from known premises.

At the time you form your belief that you are winning, you falsely believe that the sign is lit, though you correctly believe that a lit sign entails that you are winning. But, by (2), the sign is reliable evidence within your possession. This, then cannot be the case. Then, it seems, the sign must be on when you make the inference:

10) When the inference is made, it is made from true, known premises.

This does not have the problem above, but it also seems suspect. At the precise time of the inference, you turned out (at the moment you acquired your belief) not to have known that the sign had gone dark. Rather, you have acquired a false belief that you are winning. This is not consistent with (1). (1) holds that you know what the sign indicates is true. For, at the precise moment you come to believe that you are winning, the sign is dark. This follows from its reliability and its connection to your mental states.

Put another way: No doxastic attitude can be appropriate in these circumstances. Suppose that in the end, the subject ends up winning nothing. By stipulation, that means that she forms the belief that she will get the money. When she forms that belief, it is *not* the case that said belief is true for a moment and then rendered false. Since she doesn't win in this world, it has always been the case, or is at least the case once her judgment has formed. Suppose that in the end, the subject ends up winning. By stipulation, she has formed either an agnostic attitude or one of disbelief. Disbelief shares the same problems as belief. Agnosticism does not fare well either. She knows with certainty that agnosticism leads to her victory. How could she form such an attitude rationally? Strategies like regarding herself third-personally, recall, might not work; that is to say, she may not form an agnostic attitude on the basis that she has no idea where she will end up, for doing so has the perfectly foreseeable consequence of guaranteeing her winning the prize.

Thus it seems, (1)-(4) are inconsistent. It can be the case that you know the sign is lit, and that you know that a lit sign entails you are winning. And it can also be the case that you believe that you are winning, and, and that such a belief stems from your belief, just a moment ago, that the sign was lit. But it cannot be the case that you possess the evidence the sign suggests and form the belief such evidence warrants at the same time. It is always already too late for such an inference.

1.6: Two Step Views:

One might make the following objection: Of course such an inference is permissible. For, while you are making the inference, you are well within your philosophical rights to make it. It is from true, known premises. The problem only arises after the inference has installed the relevant belief, which will come into being always already false. But then that belief can be corrected by another faultless inference, which will in turn install a different, false belief, and so on. But one remains rationally unimpeachable throughout, for one always reasons on the basis of evidence, and one can't be blamed for the fact that the evidence keeps changing.

On this picture, inference (and belief-formation generally) is a two-step process. The idea is that there is an initial appraisal, followed by a subsequent instillation of a belief. Following a

venerable philosophical tradition, we can call the first thing the "judgment," and the second the "belief." On the intuitive version of this view, judgments cause beliefs to come about. And, we can suppose, even if judgments don't cause beliefs, beliefs would still follow judgments, in the sense of being temporally subsequent. If the two-step view is right, we could suppose that the sign turns off "during" the inference. Roughly, we'd want it to go dark after the reasoning is done and the judgment made, but before the belief has become installed.

If this were correct, it might undercut the timing objection and rescue the account Reaber favors. If the subject's believing P is subsequent to her judging that P, then she can take stock of the (still available) evidence that P, and make her judgment, all the while not-yet-believing P. Since she doesn't yet believe P, P is still the case, and the evidence upon which she makes her judgment has yet to disappear.

This response is both tenuous and of dubious relevance. The process here described is the subject of great controversy. Some, like Shah and Velleman (2005) seem to endorse it in a qualified way. Boyle (2011) criticizes it extensively.[5] But even if it turns out to be the correct account of belief-formation, it might not fix much. Suppose that it is correct. Is it the case that the agent is doing anything irrational? Her "judgment," after all, is correct. Now, either the agent has control over whether her judgment issues in belief, or she does not. Suppose she does. As she ponders whether to let her correct judgment issue in belief, she might recognize that it will fail to do so (for, the belief will be dead on arrival, and she knows this). As such, she might simply enjoy her correct judgment that P without believing P. But this looks exactly like a belief. She is reasoning on the basis of the truth of P, which she accepts in some meaningful sense. If she didn't meaningfully accept it, she could not reason thus. If, on the other hand, the belief-generation process is automatic, then she does look to be in a familiar kind of bind. Her judgment, while unimpeachable qua judgment, is here a false-belief-installer—and intuitively those are rationally problematic.

[5] There is slightly more extensive discussion of these questions in the following chapter.

But all of this is easily sidestepped. Even if the worries in the previous paragraph can be avoided somehow, on the picture now examined, judgment, rather than belief, looks like the real locus of rational agency. It is in judging that agents respond to the evidence, weigh up reasons, and generally do the sorts of things that we would recognize as rational or irrational in more ordinary cases. A revised paradox might simply target judgment, rather than some doxastic product further downstream. The reformulation that would do this is obvious: we simply stipulate that the thing that will be rendered false (or inappropriate, as the case may be) is judgment, rather than belief. But then the timing problem reemerges straightaway.

1.6.1: Taking Stock:

The motivation for examining this account of belief formation was to examine what happens when the faculty of human psychology that comes to a verdict about evidence (be it by judgment, or inference, or whatever else) is no longer required to do so simultaneously with the existence of evidence indicating the contrary. As such, an *inference* can take place without immediately resulting in an inappropriate doxastic stance. That doesn't happen until after the belief is "installed." But if we make the paradox concern judgment directly, the intervention will be contemporaneous with the judgment—and thereby have false content. (Content, I stress, that we could foresee would be false, because we can easily see what would be the case were we to make the inference.)

Moreover, such a move may turn out to be irrational on the pragmatic view of James and Velleman. For, whatever the merits it may have, it gives us a false belief we lacked before. Of course, I do not wish to imply that pragmatists should be the only ones suspicious of this kind of reasoning. Evidentialists too are sensitive to reasoning that leads to false conclusions, at least if we have good reason to think the reasoning will lead to a false belief. Take a standard higher evidence case: If I know that some drug has radically impaired my ability to do arithmetic, it is highly intuitive that one reason I have for not trusting the results of my arithmetic are that I know that they are unlikely to be true. *A fortiori*, I have good reason to refrain from believing that which I know will turn out false. One could perhaps argue on evidentialist grounds, then, that we ought to refrain from taking any kind of stance in this case.

It sounds almost tautological that if we reason properly on the basis of presently sound evidence, we have evidence for the conclusion drawn. Forming beliefs in SUDS cases initially seems not to work like that. This is not merely because they lead to false conclusions, for such occurrences are possible without violating any epistemic norms, since evidence is sometimes misleading. What rather seems to be the case is that one's total picture should contain the knowledge that one's evidence for some proposition, while *now* supportive of p, will cease to exist once the belief that p exists. As such, one's total evidence does not support p, even though p is true and would be available were we not so badly situated.

On Reaber's account, it has turned out, the following are claimed to be true simultaneously: First, the evidence must indicate p. Second, we must believe p. Third, p must not obtain. But the evidence cannot indicate that p and p not be the case, given that we have stipulated that the evidence is reliable. We must then revise our analysis of SUDS cases in light of this.

1.6.2: Paradoxhood, Auto-Anti-Alethic Beliefs, and Self-Irrationalizing Beliefs:

I have not presupposed any particularly firm or technical definition of paradox—nor do I have one in mind. What is more productive is to clearly identify what is remarkable about the cases involved. We began by trying to construct a case in which an agent is in a particular sort of bind. She inexorably chases Sysiphean evidence of truths she will never catch. On Reaber's initial construal, the agent is bound to do something irrational. That is to say that she has no option whatsoever that does not lead to some kind of inappropriate attitude. In the above section on timing problems I sketched an argument that agents were not bound to follow the evidence to inappropriate conclusions. Rather, they are rationally bound *not* to do so by *not* forming any sort of stance at all—*not even agnosticism*. In terms of qualifying as a full paradox, this may fail to clear the bar. But it has implications. For instance, it may show that it is not the case that there is always an appropriate stance to be taken. Or, perhaps it may mean that there is a fourth kind of stance independent of agnosticism available. I will take up that subject shortly.

While there is very little written about SUDS, there is a greater body of literature on the opposite kind of case; that is, the sort of case in which the adoption of an attitude, roughly speaking, provides a reason for holding that attitude. We saw two examples earlier in Peels (2014b) and the original James account that kicked off the whole debate. Here are two more:

Reisner (2018) coins the term "auto-alethic belief," where an auto-alethic belief is "one that secures its own truth."[6] These, he points out, are easily come by. He offers the example of the belief that I have a belief and the belief that I exist. He further offers a non-*a-priori* one loosely similar to cases we have already examined. There, a brain-scanner of a familiar sort checks your mind for beliefs about whether certain numbers will appear on a screen and (if certain other conditions are met) then displays them. The conditions are similar to mine in that, if you form an agnostic attitude or fail to form an attitude, then the number 16 will appear. Otherwise, the number on the screen will be half the number you believe will appear plus 1. So, if you believe 16 will appear, then 9 will appear. In fact, you cannot truly believe that any number will appear save 2 and -2. Both those beliefs are stable.

Reisner is interested in whether it is permissible to choose between those two beliefs. Both will come out true. As such, it seems to him that we have a genuine doxastic option: we can believe either without risking false belief—and all the other options seem bad.[7]

I am here more interested in maximally-hostile epistemic predicaments with no good options rather than ones in which we have multiple good options, so I am not interested in engaging with Reisner's cases in a robust way. However, a reader points out that Nottelmann and de Bruin[8] criticize his view in a way that may generalize to mine.

While Nottelmann and de Bruin share Reisner's broadly pragmatist convictions, they think he errs in focusing so much on truth. Better, they argue, to look at evidence. Their preferred set of beliefs, "self-rationalizing beliefs," is not the auto-alethic set of beliefs that guarantee their own truth,

[6] See especially (2018, §30.5.1)
[7] On the whole, the case is a more precise version of the original Peels (2014b) case which inspired my own variant.
[8] See (2019, esp. 259-61).

but the set of beliefs that supply their own evidence, so to speak. Slightly more precisely, a belief is self-rationalizing "whenever someone's holding it offers her evidence, which, at least *pro tanto*, constitutes sufficient reason for believing its content."[9]

There is a large amount of overlap between the sets, since we often know when a belief is auto-alethic, and, in doing so, come to possess evidence for it. But there is not total overlap. Auto-alethicism is neither necessary nor sufficient for self-rationalization. Their case for its insufficiency asks us to consider an auto-alethic brain state X "such that my belief 'I am in brain state X' suffices to bring me into brain state X."[10] Here, they argue, it might be that we have no evidence that the state is auto-alethic. Neither is auto-alethicism necessary, because it might be that we have excellent apparent evidence that a state is auto-alethic and it may nonetheless be true that our evidence is misleading.

What carries the argumentative water in the broader category of rationally self-*defending* beliefs (a label of convenience I mean to encompass all the above cases from James, Velleman, Peels, Reisner, Nottelmann and de Bruin, and many others) is not the auto-alethic quality of their cases, but the self-rationalizing quality. After all, it is sometimes perfectly rational to form false beliefs, and sometimes deeply irrational to form true ones—and the contours of rationality are what interest us. Those contours are exposed by our relationship to evidence, and only derivatively to our relationship with truth.

We might call the case I am offering "auto-anti-alethic." It is constructed such that any doxastic stance the agent takes has an inappropriate relationship to the truth. But, strictly speaking, that feature of it is less important than the agent's *evidential* misfortune. Since I am persuaded by Nottelmann and de Bruin's clarification, the important question is whether we have constructed the right kind of evidential bind. We need her to end up with a "self-*ir*rationalizing" attitude.

1.6.3: Evidence:

[9] Ibid 259.
[10] Ibid 260.

In keeping with this consideration, the precise status of the evidence in SUDS cases should be made clearer. In one sense, a lit sign indicating p (*coupled with the knowledge that it is infallible*) gives excellent evidence that p. A third party observer should certainly conform her beliefs to it. In another sense it is not evidence, in that it cannot be believed without error, and that error is foreseeable. In the same way, there is a way in which the inference is unproblematic. Ordinarily, when we know p and we know that if p then q, then we are entitled to q. Modus ponens is the cornerstone of reasoning, and it would be an obvious mistake to jettison it. But, as we have seen, something is here problematic, because p and q cannot be true at the same time, so we have a reason for thinking we are not always and everywhere entitled to the inference.

One resolution would be to argue that agents in SUDS cases lack evidence, or it may be that there is more than one sense of evidence at play. Becker (2016) suggests such an approach. There he is responding to an argument from Turri (2012) against the conjunction of "Triad," the view that belief, disbelief, and suspension are the only permissible doxastic stances, and "Optimism," the view that one of the three is always permissible. Becker argues that Turri confuses two senses of evidence in his argument.

Turri's argument, like mine, is motivated by a puzzle. In it, a group of eminent mathematicians, whose expertise and sincerity are known to you, recommend that you are not agnostic about some very obscure mathematical axiom *M*. Since you are not a mathematician, you are at a loss. You cannot be agnostic, for the evidence of their testimony is sufficient to make such a position irrational. Nor can you believe or disbelieve, for, while you grasp the meaning of the axiom, it is far beyond your abilities to suss out whether it is true. He concludes that you should then embrace another stance, one of "withdrawal," that is not identical to agnosticism. The details of his argument are not germane; the conclusion of it is, for it is similar enough to ours.

Becker thinks Turri is right that the evidence, in one sense of the term, undercuts agnosticism. In the collective sense of "should," you should not be agnostic. In the subjective sense of "should," you should be agnostic, since the total body of evidence available to you neither supports believing

the axiom nor disbelieving it. And, since the support for the axiom ought to be proportioned to your subjective evidence, and since your subjective evidence isn't decisive, you must be agnostic.

Analogously, perhaps we could extrapolate Peels' (2014b) view by saying you have conditional evidence against both q and ~q, where q is the proposition that you will win the money in a SUDS scenario, insofar as believing either would lead you astray.

Let t be the time when the timer goes off, terminating the scenario.

Let N be the proposition that you are in a winning state at t, where "winning state" refers to attitude of agnosticism or disbelief about winning, or having no attitude at all.

Let M be the proposition that you are going to win the money.

Let O be: "if N then M."

By stipulation, O is true at t, as is N. If you understand the scenario, then you know O. Further, you have all kinds of evidence for N, namely the infallible sign. But rationally believing M at t is impossible, since M is false when believed, and since you know as much.

If so, we might term the evidence for M "ungraspable" evidence. Ungraspable evidence is ungraspable in this case, only for agents in SUDS scenarios—an observer would not share the relevant blindspot. It is not in ungraspable in the sense that one cannot believe N and O, but in the sense that one cannot apply modus ponens to them and thereby infer M.

Ungraspable evidence is a species of "present evidence," which is (typically perfectly useful) evidence about what is *now* the case. Conditional evidence is evidence about what would be the case if a further fact obtains. Normally it dovetails nicely with present evidence, for normally whatever is true stays true once we believe it, and whatever is false stays false. In SUDS cases, this is not so, for the conditional evidence recommends against forming any stance (since they all come out badly), while the ungraspable evidence recommends belief, since M is true at t.

Strangely, if you know the evidence for M is ungraspable, then you know you cannot believe M, even though M is true and you know that *if* it is now ungraspable then M is true. Further, it seems

that you *do* know that M is ungraspable. If you have grappled with the case sufficiently, your evidence that M is ungraspable is part of your present evidence. The epistemic predicament thus reoccurs at a higher level in the following way: You know that the proposition "If I have ungraspable evidence for p then p" is true, and you know that you have ungraspable evidence for p, but you cannot make the relevant inference—namely, that the ungraspable proposition is true. It is a kind of blindspot that one can recognize as a blindspot, but only to a point.

1.6.4: Closure:

One might worry that this implies the falsity of closure. The closure principle, roughly speaking, states that:

9) If S knows that p, and knows that p entails q, then S knows that q.

Compare:

10) If S knows that the sign is lit, and knows that the sign's being lit entails that she is winning, then S knows that she is winning.

Indeed, our case does demonstrate the falsity of that version of closure and similar ones that make no provisions for S losing her knowledge that p while inferring q from p. In the scenario imagined, it is highly plausible that S knows that she does not believe that she is winning. It is true that she is winning, and she can know either by reflecting on her own mental state or by glancing at the imagined infallible sign that reads "WINNING." She can further know the lit sign entails that she is winning. What she cannot know is that she is winning, because it is false that she is winning once she believes that she is winning. As such, (9) and (10) are false, but not for the reasons sometimes cited. Those have to do with the subject coming to believe q for unjustified reasons having nothing to do with her knowledge that p entails q, or with her not coming to believe q ever, because she never gets around to thinking about it. Rather, like the grapes hanging inches above Tantalus' head, q seems within reach, until the moment one moves to grasp it.

It would be remarkable—and worrying—if this case showed the falsity of closure as a general principle, but it does not. More sophisticated versions of the principle like Williamsonian (2000) face no problem:

> 11) If, while knowing p, S believes q because S knows that p entails q, then S knows q.

See also Hawthorne (2005):

> 12) If one knows P and competently deduces Q from P, thereby coming to believe Q, while retaining one's knowledge that P, one comes to know that Q.

Hawthorne even addresses the worry specifically: "Why 'while retaining one's knowledge that P?' Well, one might competently deduce Q but have gotten counterevidence to P by the time one comes to believe Q."[11] Closure is an important principle, but it is unlikely to be a worry that anything said herein presents a non-cosmetic challenge to it.

1.6.5: *Paralysis*:

What about agnosticism? I have argued above that it is inappropriate, but some readers have been unpersuaded. I think there are a few more things to be said. Here is a restatement of the foundational SUDS case with simplified timing, thereby incorporating some of the lessons in previous sections.

> An agent is implanted with a brain-scanning device which accurately reports her mental states to a trustworthy neuroscientist. It scans her brain continuously, He promises her an immediate[12] $1000 if and only if she believes that he will *not* give her $1000 *or* if she forms an agnostic attitude. If instead, she forms the belief that she *will* receive the money, the experiment terminates and she leaves with nothing. If she forms no stance, nothing happens.

[11] (2005, 43)

[12] Removing the timer anticipates an object along the lines of Nottelmann and de Bruin's (2019) worry that agents in such systems might have a right to agnosticism given that it's very difficult to know exactly what one will be thinking in the future in such a fluid scenario. Without a timer, we have less to worry about.

In *Paralysis*, the agent enters into a situation with several epistemically bad options. If she forms the belief that she will get the money, she will form a false belief. If she infers that she will not get the money, she also forms a false belief. These, as we have seen, are problematic. This, I have argued, is because her evidence is ungraspable.

What if she forms an agnostic attitude? This is less bad in one way: she does not form a false belief. However, it seems to me that the same inappropriateness that underlies belief and disbelief underlies agnosticism. If agnosticism is just the state of having a middling credence, then her credence would miss the mark widely.

Suppose it is rationally obligatory for her to be agnostic. it would not make sense for her, at the moment she forms her attitude, to limit herself to betting only $500 that she will win the money (she should be willing to bet whatever she has). This is not because she knows she will win, but because she knows the following:

13) She is in a SUDS scenario.

14) People in this kind of SUDS scenario win money if they form agnostic attitudes.

15) She has an agnostic attitude.

16) If the previous are satisfied, it is a good idea to behave as if one will win the money.

Thus, we arrive at the surprising conclusion that she should bet the farm. (13) and (14) are known by stipulation. We have already discussed why it seems plausible that she knows (15). (16) is easily knowable. All she has to do is think about the scenario from a third-person standpoint. This betting pattern is remarkable. For one, it looks like the ungraspable evidence is not so useless if it can be action guiding. Why should the rationality of such a bet be available to the agent if agnosticism is the right policy?

As Friedman (2013) argues, and an agent is agnostic about p when she has p "on her research agenda;" that is, when the agent has the right dispositions to be sensitive to new information that might help settle whether p, then we have further reason to think she shouldn't be agnostic about p.

The agent in question knows that she will never match her credence to the truth, and thus that there is no point in trying—better to save the brain space for contemplating something else.

It is admittedly true that if she knows that she cannot rationally infer either p or ~p, then she will never know either proposition. If knowing that you will not know either proposition is sufficient justification for agnosticism, she has an easy (and perhaps epistemically obligatory) route to her $1000. And if Friedman is right about agnosticism also being an attitude of committed neutrality, it seems that knowing one will never know whether p might well count as agnosticism.

However, an agent would be able to know that agnosticism *would be* inappropriate, were she to be agnostic, given that it mismatches credences with results. Further, she can know that she is agnostic when she is agnostic. She can thereby infer that it is inappropriate. That it is inappropriate, notice, is not ungraspable the way believing that she will win is ungraspable. Permitting agnosticism seems to give undue privileges to middling credence for no sufficiently persuasive reason.

Arguably, agnosticism is then inappropriate. But if I am wrong, it is not the case that all this has been in vain for reasons we have already seen. Namely, that a rational agent in a maximally epistemically-vexed case like the above would win the prize money every time, no matter how many iterations she goes through.

If, conversely, I am not wrong, then all three attitudes are inappropriate.

1.7: Upshots:

There are significant upshots to the conclusion that she must refuse any kind of attitude. First, because such a refusal implies the falsity of a commonly held (and intuitively appealing) epistemic principle:

> Optimism: For any set of evidence E and for any proposition Q, E will, all things considered, support either believing Q, disbelieving Q, or withholding judgment[13] on Q. (Turri 356).[14]

SUDS-vexed agents have what Sorenson (1988) calls a blindspot, for they cannot see the truth of a proposition because of their vantage points.

For those committed both to the Ramseyan view that doxastic states are credences, and to optimism, other options remain. One could maintain that agents are permitted to embrace attitudes they foresee will be inappropriate, despite the foregoing, but this, I take it, would a very undesirable result.

Alternatively, one might think that refusing to believe, disbelieve, or be agnostic itself constitutes a kind of stance. Turri (2012) suggests there might be a stance which he calls "withdrawal." If this is right, one can preserve uniqueness, optimism, and the credence framework, but at the expense of:

> Triad: There are only three doxastic attitudes—belief, disbelief, and withholding—and once you've [sufficiently] considered a proposition, there are, intellectually speaking, only three options open to you: you either believe it, disbelieve it, or withhold judgment. (355).[15]

Since this too is a widely held[16] tenet, this approach is also quite revisionist. We can therefore conclude that some popular thesis of epistemology is false, albeit without knowing which one.

Two readers commented that they did not think Triad obviously popular, since it is demanding. If they are correct, and they may be, then we still have a new reason to reject it. However, there does seem to be something *distinctive* about *Paralysis* and similar SUDS. *Paralysis* is a scenario

---

[13] I prefer "being agnostic about Q" to "withholding judgment on Q," but the difference is verbal, I believe. In case it is not merely verbal: I mean the principle to refer to Friedman-style agnosticism, where "agnosticism" refers to a substantive stance like the one discussed above.

[14] This too is widely held. See Conee and Feldman (2004: ch. 4) and White (2005: 445-59).

[15] See also Sosa (1991), Chisholm (1989), Feldman (2002), Bergman (2005), Pace (2010), Ryan (2010) for a few examples of endorsements (sometimes implicit) of Triad.

[16] Smart (2020) endorses Triad, but argues that it is not popular. However, that with which he takes issue is the claim that philosophers think there is a substantive difference between belief and disbelief. Smart thinks there *is* a substantive difference, but that most philosophes think that disbelief in p just is believing that not-p. If he is right, then the more popular view might be called "dyad." However, nothing I have said concerns the distinctiveness of disbelief; I am arguing that possibly there are too *few* options, not too many.

in which a millennium of deliberating over what is rational to believe would not (rationally) lead to any of the three stances. That a scenario does not lead to belief or disbelief is unremarkable. There are an infinite number of questions the answers to which can never be obtained by agents like us. But for those, upon enough reflection, we are almost always entitled to throw up our hands and become agnostic on the question. In SUDS, agnosticism is *in principle* inappropriate *regardless* of how long we think about it. The only attitude, if it is an attitude, that is rationally permissible is a complete abandonment.

1.8: Psychological Observations:

All I have said so far concerns the question of what is rational to do in SUDS cases. But there is also the matter of what is psychologically possible for agents like us. It might be the case that it is rational for the agent to refrain from making inferences about what is true in these cases, but it seems to me that no one would be able to. This goes, I think, for any approach. Suppose that my own hunch is correct, and we are rationally obliged to refrain from forming an attitude. In *Paralysis*, it strikes me as impossible for an agent to refrain from believing that she is winning when she knows perfectly well that she has no stance (perhaps a sign tells her this if introspection cannot), and knows that not having a stance entails, in this instance, that she will never get the money. I am not, I emphasize, arguing that she should form this belief—merely that she would. Reflecting on these cases shows as much. A more vivid one might drive the point home:

Socrates is committed to maintaining his epistemic purity above all else. His nemesis has captured him and connected him to a device which very slowly pumps Hemlock into his bloodstream. Socrates knows how this diabolical device works: Should he form the belief that Hemlock is being pumped into his blood, the device will deactivate at exactly the time his belief comes into existence, such that if he ever believes that it is working, he believes something false. He knows all of this.

If Socrates is anything like me, he would lose his epistemic purity the instant he wonders whether Hemlock is being pumped into his blood. For, when he wonders this, it is true that Hemlock is being pumped into his blood, and in an apparently manifest way. Even if he knows believing this

would be false, it seems inescapable that he could wonder "Am I being poisoned?" without adverting to the fact that he is being poisoned. As soon as he has done this, he has lost, and will have to content himself with living an epistemically sullied life.

This is not to say that we cannot distract ourselves in such situations. But I take it that this is a very uninteresting way in which we can get around these problems. If we bracket it (perhaps some threat makes it morally obligatory for Socrates to think about his situation), then we are left with the ensuing epistemic failure. If he were unlike us, then he would be able to refrain from ever forming a false belief, and die of Hemlock poisoning. Note that he would indeed have to die—for even at the last moment of his life, it would be the case that his believing that he was currently being poisoned would be false, were he to believe it. Of course, this would hold true regardless of the practical concerns (e.g. if money were flowing into his bank account instead of poison into his veins). We are weak like this, I think, regardless of the stakes.

The way in which we cannot avoid thinking about what is now the case, even if we know it is only ephemerally so, may give us some reason to think James and Velleman allow for too much. If I am promised that you will give me ten dollars if and only if I can but believe, right now, that you will, I think I would no better off (epistemically, anyway) than Socrates. For, as soon as I wonder what is now true, I cannot help but see that it is not now true that you will give me ten dollars, and the game is over.

To sum up: SUDS cases involve truth-makers concerning one's own doxastic states. If we wonder about the outcome of the cases, we realize that in order to know what will happen, we have to know what we think. It is not a necessary truth that we can only do that by asking ourselves whether p. There is nothing incoherent about a sign like the one referred to earlier. In the ordinary case, we advert to the world in order to know our own view.[17] We know that our view is our view by seeing ourselves think it, or perhaps just by thinking it. In the latter case, we still advert to the world—in this case, by looking at a sign in the knowledge of what the sign entails about us. In every case, then, we

[17] See Evans (1982).

know what we think by looking at the world and seeing what is true. Not what would be true were we to embrace it—but what is true. This is why SUDS cases are frustrating. We cannot know without looking at what is actually the case, and doing that is what gets us into trouble.

1.9: Conclusion:

When I first began looking into these cases, I did so with an eye to undermining the Jamesian position. It seemed to me that if we allow for rational agents to simply jump to "safe" positions akin to the Jamesian one about victory in a footrace then we allow for too much. I still think as much, but I do not take myself to have shown it here.

Many refinements I have not discussed are possible. For one, Pragmatism could be combined with a robust doxastic consequentialism and, with respect to SUDS, take a line whereby any stance is permissible because any stance is equally bad. Doing so would perhaps preserve at least Triad. Further objections are also possible. It may be that I have assumed there is such a thing as simultaneous causation, which is a metaphysically controversial stance.

I have argued that SUDS cases show that something is amiss in the conjunction of Triad and Optimism. That is to say that it is not the case that one of belief, disbelief, and agnosticism can always be appropriately embraced. This is remarkable, though it will not end any debates between pragmatists and evidentialists. On that matter we must remain agnostic, or perhaps stanceless, if such a non-position is possible.

Chapter 2: Recent Unanswered Critiques of Volitionism:

2.1: Introduction:

In the previous chapter I examined a possible exception to the rule that we lack direct voluntary control over our beliefs. Regardless, many philosophers hold that *ordinarily* we lack that control. This introductory section concerns where that leaves moral responsibility. I will not examine arguments for any particular a definition of "belief;" instead, I will look at arguments regarding whether and how belief is under our voluntary control, and how we might be responsible for it, beginning with William Alston's argument that we are not usually responsible for our beliefs. In what follows, I will defend Alston's conclusion from two prominent objections that have yet to be addressed in more than a cursory fashion in the literature.

In his "Concepts of Epistemic Justification" (1985), and his *Epistemic Justification* (1989), Alston denies that epistemic justification should be understood in terms of obligations that might ground praise or blame. On the picture he rejects, I am justified in believing *p* if and only if I do not violate any obligations with respect to believing, and I am therein directly responsible, e.g. praiseworthy or blameless. But, for Alston, I do not have these obligations because I do not have direct voluntary control over what I believe, and I do not have this kind of control because I cannot decide to believe *p* or refrain from believing *p*. Being able to decide to believe and being able to decide not to believe are jointly necessary conditions for direct voluntary control. I cannot change my beliefs by intending nor can I change them "at will," so I lack control over my beliefs.[18] Since the argument appears valid, if Alston is wrong and we can properly be praised or blamed for our beliefs, then one of the following four premises must be wrong:

1) We are not morally responsible for our beliefs unless we have obligations to believe.

[18] I am ignoring some details here. Alston distinguishes "basic doxastic control" from "immediate doxastic control." When someone has basic doxastic control, she can change her beliefs "at will," in the same way she can raise her arm (1989: 122). When someone has immediate doxastic control, she can change her beliefs through an uninterrupted intentional act, in the same way she can flip a switch (1989, 129). These two kinds of control jointly compose direct, voluntary control. Nothing important rides on the distinction for our purposes.

2) We do not have obligations to believe (or disbelieve) unless we have sufficient control over our beliefs.

3) Normally, the only relevant kind of sufficient control is direct voluntary control.

4) In normal circumstances, we do not have direct voluntary control over our beliefs.[19]

5) So, we are not normally directly morally responsible for our beliefs.

I will refer to (5) as "doxastic incompatibilism."

2.2: Conceptual Groundwork:

I will focus on the objections to (3) in the contemporary literature. But I will first clarify some terms. Now, (3) is the contention that only direct voluntary control suffices to ground direct moral responsibility. I have already partially distinguished direct voluntary control from indirect voluntary control. If some state of affairs is under my direct voluntary control, it is within my power to bring it about without further mediating actions. If I am in a dark room, I do not now have direct voluntary control over whether I believe the room is lit. I cannot believe that at will. However, as long as I have the ability to turn on the light, my doxastic attitude is within my indirect voluntary control; all I need to do is go flip the switch, and I thereby produce the belief that the room is lit.

The direct/indirect distinction with respect to moral responsibility runs along similar lines. If I am morally responsible for something, I am *directly* responsible for it just in case I am morally responsible for it, and *not* in virtue of being morally responsible for something else. If I were morally responsible for some φ only in virtue of being morally responsible for some ψ, then I would be *indirectly* morally responsible for it. The most commonly cited example of this distinction concerns drunk driving. Generally, drunk drivers are responsible for the damage they do, but only indirectly. The drunk driver, *given her drunkenness* and *given that she decided to drive*, could not have been

[19] Contrary to what many seem to think, Alston never actually claims that we *never* have direct voluntary control; only that we rarely do, at best (1985: 64; 1989: 136).

expected to avoid doing the damage she did. As such, many hold that she can be excused from any direct responsibility for crashing the car. Of course, she is still blameworthy for that in virtue of being *indirectly* responsible for it. And she is indirectly responsible for crashing the car by virtue of being directly responsible for getting drunk and perhaps for deciding to drive. Similarly, if I lazily leave a trap door open which I am obliged to shut, and someone falls through it and injures herself, I am not *directly* responsible for her injury (as I would be if I pushed her), but indirectly.

It is very commonly held that we are sometimes morally responsible (henceforth, just "responsible") for our doxastic states. If David, a medical student, chooses to text his friend instead of paying attention to a lecture (knowing that he risks missing life-saving information in doing so), he is blameworthy for his ignorance when it later prevents him from doing his job effectively. In such a case, David is indirectly blameworthy for his ignorance. He is directly responsible for not paying attention.

Many philosophers, however, think one can be directly responsible for beliefs. This does not exclude one's *also* being indirectly responsible for the same beliefs. For example, if I believe that justice is the advantage of the stronger, I may be blameworthy for that belief in virtue of my lazy intellectual habits, my knowingly choosing the company of vicious souls, etc. *and also* blameworthy just in virtue of my believing something repugnant. On this view, we can be blameworthy for wicked beliefs *even if* they are *not* the product of blameworthy acts. It may be that someone's heinously racist beliefs are merely the product of her upbringing, and not the result of a knowing omission to think carefully and read widely, or choosing to keep bad company, and so on. Even so, according to some philosophers, she is blameworthy for them.

For our purposes, the voluntary is what is willed, or perhaps chosen, or perhaps intended, or perhaps decided. I think the details here are difficult to work out, but a general sense is enough for my purposes in this chapter. I emphasize in passing that I am *not* making any claims about physical determinism. Most philosophers think that willing or choosing or intending or deciding are compatible with determinism; some disagree. I am not interested in this question here. For my part, I

do embrace some version of the Principle of Alternative Possibilities ("PAP")—the view that whatever ultimately grounds moral responsibility requires some kind of ability to do (or choose, or decide, etc.) otherwise. Nonetheless, since this view may beg the question against my opponents, I will not assume it here. Argumentatively-speaking, my intent in this chapter is purely negative. I aim not to positively prove doxastic incompatibilism (or the PAP-infused volitionism that almost always undergirds it), but merely to show that two very novel and very clever arguments against it fail.

Now, the most numerous opponents of doxastic incompatibilists are called *attributivists*[20], or sometimes *attributionists*. "Attributivism," as I am understanding it, is an umbrella term for those who reject *volitionism*, where volitionism is the view that some variety of voluntary control is required for direct moral responsibility. Volitionism implies doxastic incompatibilism, and extends similar principles to other attitudes and states, since we lack direct voluntary control over our e.g. some of our emotions in the much the same way we lack control over our beliefs. Attributivists reject this on a variety of grounds, but typically agree that a core problem for volitionism is that it is unable to account for cases in which we are morally responsible. Surely, they say, the Nazi is culpable for her embrace of Nazism even if she never decided to embrace it in the doxastic sense. Volitionists, for their part, usually admit that there are cases of agents believing morally heinous things, or having repugnant attitudes, or perhaps unintentionally doing wicked things, without directly responsibility, since these horrors are not the result of an exercise of those agents' voluntary control. But the volitionist can try to defend herself against the charge of being too easy on the wicked, for she can argue that such wickedness can be *traced* to exercises of voluntary control. It is on these tracing questions that much of the contemporary discourse focuses.[21] Can choice really account for all that is

[20] Some prefer simply "non-volitionists," because there is widespread disagreement among these theorists about whether the *mere* attributability of a given attitude to a given subject is sufficient grounding for blameworthiness. See Fritz (2018). However, I am not here interested in internecine debates within this family of views except insofar as they may help clarify a specific objection to volitionism.

[21] See, for example, Vargas (2005), Fischer and Tognazzini (2009), Timpe (2011) and Shabo (2015). As an aside: One way to get a very quick grasp over the fundamental disagreement is to note that tracing debates go two ways. Everyone agrees that our choices have effects on our character; equally, our character obviously bears on the sorts of choices we make. The question, at its most basic, is about which is primary: Are we blameworthy for being bad because of the bad things we have done, or are we blameworthy for the bad things we have done because we are bad? Volitionists think choices lie at the very bottom; attributivists think it ultimately comes down to character.

usually blameworthy in character? In belief? How does this tracing go? Where can it be blocked? This, at any rate, is a distillation of the central disagreement—though it overlooks instances where there seems to be real disagreement about the verdicts we ought to take with respect to specific cases, rather than just disagreement about the explanations for the justifications of those verdicts. For simplicity's sake, I will confine discussion of cases to those underdiscussed ones upon which much of this chapter is based.

Attributivists come in many varieties,[22] none of which will receive adequate justice here. Some think that moral responsibility is fundamentally a question of one's identity or commitments.[23] For some, attributability is crucially distinguishable from accountability, where accountability concerns an agent's violations of duties (as opposed to the mere possession of a condemnable attitude), and where only accountability is sufficient grounds for blame.[24] The account we are about to examine is an "answerability" account of the sort prominently defended by Angela Smith (2015), and in the tradition of Anscombe (1957). An agent is answerable for some action or attitude if and only if the question "Why did you do that?" is not refused by the agent saying something along the lines of "Oh, I wasn't aware I was doing that!" (Or, if I was aware, then the state expressed is not mine in a sufficiently deep way.) If the agent is able to answer, her answer will appeal to her reasons and reveal her character, and in virtue of the answer she will be subject to the reactive attitudes generally taken to be characteristic of moral responsibility. It is to one such answerability view that we now turn.

2.3: The First Unanswered Challenge:

On Pamela Hieronymi's attributivist view (2006, 2008, 2014), we are most fundamentally responsible for some X insofar as X "reveals our take on the world and our place within it—it reveals what we find to be true or valuable or important" (2014, 19). Literal revelation is not at issue here. We are responsible for what we do whether or not our doing it happens to reveal anything to anyone. What matters is what we take to be true or valuable or important, which is expressed in how we

[22] I am indebted to Fritz (2018) for an excellent recent general taxonomy of attributivist and volitionist views.
[23] See Watson (2004) for a particularly influential account.
[24] See McKenna (2012) for a critique of the way this distinction works in Watson. Further note that if this analysis of attributivism is correct, then it is strictly compatible with volitionism.

implicitly answer questions about what is true (by believing), as well as how we answer questions about what to do (by acting). Why are we responsible for these things? Because we have a special kind of control over them, what she calls "evaluative control." This fundamental kind of control is roughly mental, at least insofar as it lies in our mental takings. Here is her own description:

> If an attitude embodies our answer to a question or set of questions, then it seems we will form or revise such an attitude in forming or revising our answers to the relevant question(s). If you become convinced that p, and so settle for yourself the question of whether p, you therein, *ipso facto*, believe p. Likewise, if you settle (positively) the question of whether to φ, you therein, *ipso facto*, intend to φ. Moreover, if you change your mind about whether to φ, or about whether p, in such a way that you are no longer committed to φ-ing or to the truth of p, then you no longer intend to φ or believe that p. We might say that we control these aspects of our minds because, as we change our mind, our mind changes—as we form or revise our take on things, we form or revise our attitudes (2008: 372).

The upshot here is that we control our minds by thinking with them. We are rational beings, and our core takings about the world are governed by reason that, far from restricting us, is the very source of our freedom.

This kind of account can explain a great deal about the scope of moral responsibility. On most philosophers' pictures, we are responsible for what we freely do, that which is under what Hieronymi calls "managerial control," or "discretion." Discretion involves settling something by acting on it. I sometimes have discretion over whom I vote for, what I eat, what I study, what I think about, etc. So, importantly, I have discretionary control over the world around me and (to some degree) over my mind, the very source of evaluative control. This is because I can do things to the world and my mind *for any reason I take makes doing so worthwhile*. Insofar as this is true of an action, that action is within my discretionary or managerial control.

In the above quotation Hieronymi lays out how we are responsible for the domain of evaluative control. But how are we responsible for what we "manage" through discretion? In brief, we are responsible because our discretionary control ultimately bottoms out in evaluative control.[25] When

[25] Is this true? Hieronymi assumes as much, but this would imply that akratic acts (if such things exist) do not, or do not fully, bottom out in our evaluations. Very plausibly, we sometimes act against what we take to be right. Hieronymi may respond by rejecting that akratic acts exist, or she may think that, when they do occur, they constitute a kind of change in evaluation. I think the first approach is more compatible with her general line

I vote, what settles my intentional action is, for the most part, what I take to be the relevant considerations in the election. It is likewise with even those actions which have doxastic consequences. Perhaps I decide to reflect on my theological views because they have been challenged; suppose on doing so, I change them. This is an instance of our managing something through our discretion (by deciding to reflect), and thereby indirectly changing our evaluative apparatus (through the subsequent change in belief). But notice, however, that what pushes the managerial decision to reflect was a further act of evaluation: I took the theological challenge as worthy of answering—so much so, in fact, that I sought to answer it. As such, even in this case of managerial activity altering my evaluative apparatus, we discover, at bottom, more evaluation.[26]

When I freely choose, in paradigm cases, what makes me the proper author of that choice, what explains why I am to be held responsible for it, is often left unspecified by my volitionist allies.[27] *Ceteris paribus*, Hieronymi's account goes a bit further. For Hieronymi, we can control the very minds that evaluate the various goods on tap just by evaluating those goods. There is no need for extra managerial faculties, at least not ultimately.

What is the volitionist to make of this? In the final account, she must reject the view. Volitionism has no precise definition (for all I have said, anyway), but the defining characteristic of the view, as I understand it, is that choice, not evaluation, is the source of responsibility.

However, there is a feature the volitionist might find attractive, if only initially. Hieronymi is here offering a sense of control which we seemingly exercise over our actions, but not over our beliefs. I have discretion over my actions just in case I can perform them for any reason I take makes performing them worthwhile. This is initially plausible. I may bring my tired friend a cup of tea, or throw it in her face, *for any reason I take makes doing so worthwhile*. Since these actions are of the sort that volitionists usually think ground direct responsibility, perhaps her notion of discretion is an

---

of argument. However, since this does not directly pertain to my main objection to her argument, I won't pursue it further.

[26] Note that this seems to be false in some domains. We do not always, for instance, think about things we think are worth thinking about. Nonetheless, it is initially plausible for other domains.

[27] Although, see Shepherd (2015) for a survey of volitionist-friendly accounts of the control we exercise over our decisions.

apt candidate for the kind of control volitionists care about. And indeed, it passes other volitionist-friendly tests. Importantly for our purposes, we seem to lack discretion over our beliefs. If I could spend the next half hour in the thrilling belief that I am great Pharaoh of all Egypt, I certainly would; lamentably, I am a mere slave to the evidence as it appears to me. I cannot believe, or disbelieve, or suspend judgment on something, for any reason I think makes *believing* or *disbelieving* or *suspending* worthwhile. "Discretion" gives the volitionist verdict on action and belief, so perhaps it can be of use.

This is not an accident. Hieronymi thinks volitionists get something right; namely, that there is a kind of control we have over our actions that we do not have over e.g. our beliefs. It is an important kind of control for her, and she finds volitionist intuitions are not completely unfounded. However, she observes, what we are calling "discretion"—the ability to φ *for any reason we take makes* φ*-ing worthwhile*—doesn't seem to rule over our intentions, for which we are surely morally responsible. This seemingly volitionist-friendly kind of control and moral responsibility thus come apart. For, when I intend to make a cup of tea, I do not do so because I think *intending* to make a cup of tea is worthwhile. I intend to make a cup of tea because I think *making a cup of tea* is worthwhile. It might even be true that I have a reason to intend to make a cup of tea, since my intending to make one usually leads to my making one. But that is not normally why I intend to make a cup of tea. I intend to make a cup of tea because I want to make a cup of tea, not because I want to intend to. At a minimum, it is not the case when our motivation to intend and our motivation to act in the way intended are at odds.

Hieronymi supplies a well-known example. In Kavka's (1983) toxin puzzle, an eccentric millionaire promises you a million dollars if you intend, right now, to drink a certain toxin when it is presented to you by his henchman tonight at midnight. This toxin, while non-lethal, will make you very ill for a short period of time. Actually drinking the toxin has no bearing on whether you get the money; the money is *solely* tied to your now intending. Seemingly, you have every reason to form the intention, but your efforts to do so will fail, for you will realize that you have no reason to actually drink the toxin later. So, we lack discretion over our intentions. We cannot intend for any reason we think makes intending worthwhile. Rather, we can only intend for reasons we think make acting

worthwhile. Since, she argues, we are presumably directly morally responsible for our intentions, discretion does not overlap with the ultimate locus of moral responsibility.

2.3.1: Problems with Discretion:

In the following I argue that "discretion" as Hieronymi defines it fails to overlap sufficiently with what volitionists are after to be even initially persuasive to them. Other versions of the view not only do better than what we might call "discretionism," but they actually make more sense of toxin cases than Hieronymi's own volitionist view.

First, I examine some minor problems with the putative relationship between discretion, action, and intention. We sometimes lack discretion over our actions and sometimes have it over our intentions. Hieronymi's account does not carve at the joints to the extent we might initially think. I do not think these quibbles amount to a devastating objection, but they increase the motivation for an alternative view that better accommodates intuitions about the categories of human activity for which we are responsible.

Next, I argue that volitionists ought to be careful about leaning heavily on "action" accounts of decisions in order to defend themselves from the objection that volitionism cannot provide for moral responsibility over intentions. Nonetheless, that path is appealing to some, and thus may be of philosophical significance.

Finally, I argue that Hieronymi's own view loses intuitive force when we look more closely at "toxin-style" cases. Normally, when we have sufficient reason to do something, we can straightaway decide to do it. Our reasons for doing it do not come apart from our reasons for intending. In toxin-style cases, those sets of reasons do come apart, and we are unable to form intentions for reasons we think make doing so worthwhile. On Hieronymi's view, we are responsible for intentions tout-court. Or, if that's too strong, she suggests we are at least sometimes responsible for forming (or failing to form) intentions in toxin-style cases. But an examination of these cases shows her view to be highly unintuitive. Both Hieronymi's view and traditional volitionist approaches agree that we are normally

responsible for intentions, but traditional volitionist views fare better in toxin-style cases, as does, for that matter, her own candidate form of volitionism

2.3.2: Illuminating Quibbles:

In the following, I argue that we *do* sometimes exercise discretion over our intentions, that we frequently *do not* exercise discretion over our actions, and that there seems to be no close match-up between discretionary control and either the category "intention" or the kind of control volitionists are interested in. This does not imply that the kind of control she discusses is incoherent, but close examination helps reveal its orthogonality. It would be one thing if her suggested version of volitionism failed only to capture the desired result with respect to responsibility for beliefs; however, it also fails to cut the categories of intention at the joints. Second, it has the same problem with respect to action. There, Hiernoymi-style discretionism plausibly fails to ascribe responsibility to a class of cases for which we are responsible. As such, the volitionist has some motivation to think the failure of discretionism particularly threatening.

It is not a problem for Hieronymi that discretionism fails to delineate accurately the sphere of moral responsibility. That is her own thesis, after all. The problem is rather two-fold: First it fails badly enough in enough places that it will not even be tempting to volitionists. If, for instance, it implied only that there was a weak sense according to which we are directly responsible for our beliefs, but were otherwise attractive, it might change volitionist minds about doxastic incompatibilism. Second, there are simply better alternatives available, one of which I will partly elaborate in a further section.

I examine the discretion-intention connection first. In a little-noticed paper, Clarke (2008) argues that we sometimes *do* have discretion over our intentions. He submits the case of Ann, who is trying to get work done, but finds herself distracted by the question of whether to go to a certain meeting. She wants to settle the matter so she can get back to work. She doesn't like meetings in general, but she's a bit curious about this particular one. Now, she knows her friend Sam is curious about what she would say at this meeting, and that he wants to know whether she will go (so he will

know whether to come and hear her observations). Ann correctly supposes that she has equal reasons for and against going to the meeting, and she also correctly supposes that if she decides to go, she will stick with her decision, enabling her to stop deliberating and get back to work. If she decides not to, she will be distracted by thoughts of the meeting. Ann decides to go, and tells Sam she will go; when the time comes, she goes. "It seems that one reason she has for so deciding, and a reason that she has for subsequently intending, is that the decision and subsequent intending are more likely than the alternatives to promote her getting work done between now and meeting time. But this consideration is not a reason to attend the meeting" (196). Rather, in part, she intends to go to the meeting because she thinks *that intending* is worthwhile.

There are objections to this line of reasoning. Perhaps, in this case, Ann merely has reason to tell Sam that she will go to the meeting, and thereby receives reason to attend it. Here, her reason for intending to attend the meeting is not that she thought intending to do so was worthwhile, but only that attending was worthwhile (she doesn't want to disappoint Sam). This is the sort of line suggested by Audi (1993). On this picture, Ann does not have a reason to intend, but rather a reason to cause herself to intend (in this case, by promising Sam). Analogously, Audi might say that a belief's being beneficial is not a reason to have the belief, but a reason to bring about one's having it.

Taking this line risks ignoring Ann's reasons. On Audi's picture, since only an intention to attend the meeting will put the matter to bed and allow her to get on with her work, Ann has reason to intend to intend to go to the meeting. That is, since she has reason to bring it about that she intends, and she can most effectively bring it about that she intends by deciding to intend to go to the meeting, she has reason to form the intention to intend to go. But, as Clarke implies, if she stopped after forming the meta-intention (perhaps she is distracted by an electric shock), she would not have responded to the reasons she has.

Ann does not just have a reason to intend to intend to go. Nor does she merely have a reason to decide one way or another (which, at the risk of stating the obvious, is the ordinary way of forming intentions). "As the case is imagined," writes Clarke, "what Ann does is decide to attend the meeting. The decision doesn't cause Ann to come to have the intention in question, and the decision doesn't

simply bring it about that she so intends; rather, the decision *is* her actively coming to have that intention, it is her active formation of that intention" (201). No daylight is visible between our having reason to intend to do some particular thing and our having reason to bring it about by deciding to do it. But then it looks like the obvious candidate for what could count as our causing ourselves to have a certain intention, namely our deciding, is not always already a reason that makes intending worthwhile.

Plausibly, then, we *do* sometimes have discretion over our intentions.[28] Ann intends to go to the meeting, at least in part, for a reason she thinks makes intending to go worthwhile. Of course, once she tells Sam, she *also* has a reason to intend because she thinks going to the meeting is worthwhile (that is, in addition to a reason to intend to go). Nevertheless, her intention passes the discretion test: she intends because she thinks intending itself is worthwhile.

I have no intuition whatever why Ann would be any more or less morally responsible, nor more or less directly (or indirectly, for all I have argued) morally responsible for this sort of intention than for any other. Her industry seems just as praiseworthy without the discretion.

A volitionist is most likely to think good candidates for praise or blame have no particular relation to the cases of intention over which we enjoy discretion. I may be directly responsible for something over which I have discretion, or I may not; I may be blameless for something over which I lack discretion, or I may not. The division Hieronymi makes simply overlaps little with the way we pre-theoretically understand moral responsibility.

Just as it is dubious whether we lack discretion over our intentions, so it is unlikely that we have always discretion over our actions. Recall that Hieronymi thinks we have discretion over actions insofar as we can perform them for any reason we think makes doing so worthwhile. This is true of ordinary actions. I can call my mother or open the window for whatever reason I think makes doing those things worthy. But there do seem to be limitations, like classic Davidson-style cases (1971, 50).

[28] Admittedly, this is only so in cases where our reasons are fairly balanced.

In Davidson's canonical case, he imagines a man who is skeptical that he can make ten carbon copies. He tries anyway, and, surprisingly, he is able to make the copies.

In this case, the agent does *not* make the copies for a reason he thinks makes doing so worthwhile. Rather, he makes the copies because he thinks there is a reason he thinks makes *trying to do so* worthwhile. By stipulation, he does not believe he is making the copies, so it would be strange if his action were *necessarily* accompanied by a motivating belief about the worthiness of doing it. His ability to make the copies for a reason is, in some sense, blocked by his belief about what he can do.

In Davidson's original case, nothing moral hangs in the balance, but we can easily imagine cases where it might. In such cases, the agents in question plausibly act responsibly/culpably. If Gretchen is skeptical of her ability to shoot an innocent person, but gives it a try and turn out to be very able to do so, she has done something wrong and is culpable for it. Yet, doing what she did was not something she did through her discretion.

None of this is uncontroversial. Davidson's case is a classic wrench in the works of post-Anscombe philosophy of action, and there is no obvious way to accommodate it while preserving the clear epistemic dimension of intentional action accepted by many philosophers. It is no particular mark against Hieronymi's idealized volitionist (that is, her imagined defender of discretion as the basis of responsibility) if she doesn't accommodate the objection. Here's one way she might: She might suggest that Gretchen is not directly responsible for shooting, but is directly responsible for trying to shoot. After all, she did *try* to shoot for a reason she thought was worthwhile.

I have no strong view about the success of this line of argument, but some might. It does seem that Gretchen shot someone intentionally, and it seems odd to say she is not directly blameworthy for doing as much, *ceteris paribus*. Nonetheless, it is another cost of embracing discretion-based-volitionism.

2.3.3: Alternatives to Discretion:

What else might the volitionist embrace? Here is an attempt from Peels (2017), in response to Hieronymi's (2006) argument[29] that volitionist's cannot exclude direct responsibility for belief without also excluding direct responsibility for intentions:

> [One] could say that one has control over ϕ-ing and one is responsible for ϕ-ing only if ϕ-ing, in the situation at hand, can be the intended product of our uncoerced will or the exercise of our uncoerced will itself. Thus, many of our actions are under our control because they are the uncoerced products of our will: we perform them because we *will* to do them and are not coerced in willing to do so. For actions, this means that we form an uncoerced intention to perform the action. Intentions themselves are usually also free and we are usually responsible for them, because they are are the exercise of our uncoerced will… Beliefs, however, are neither the product of our uncoerced will nor the exercise of our uncoerced will: we do not believe as the result of willing to believe something, nor is believing something identical to willing something without being coerced to do so (78-9).[30]

Something along these lines will turn out not only to be more defensible than Hieronymi's candidate form of volitionism, but also more defensible than her actual position. We can call the view *broad* volitionism ("BV"), since it captures volitionism-friendly intuitions about a broad variety of cases, including intention and belief: It gives the volitionist responsibility for intentions, albeit without settling questions about the directness of the responsibility involved. However, it does so without accidentally rendering us responsible for our beliefs in the way Hieronymi's imagined "discretion" view does. Discretionism gets intention wrong because it is oddly gerrymandered. Discretionism holds that only Clarke-style intentions (see the Ann case in 2.3.2) are strongly controlled enough to bear responsibility. Second, it can deal with Velleman-style cases without courting controversy, for it is the case that Gretchen's act of shooting was intended product of her uncoerced will if Gretchen shot intentionally.

Whether this version of Volitionism gives a more convincing verdict on the question of belief will depend on the commitments of the reader. But I am not primarily arguing that Hieronymi's view is untenable. I am rather arguing that her discretionism is escapable; that there are other versions of volitionism that do better in a number of regards. This one is not so unnaturally gerrymandered, nor does it give the wrong verdict repeatedly.

---

[29] Owens (2008, 81) makes a very similar argument to Hieronymi's at an earlier date.
[30] Another voilitionist, Booth (2009a, 9-11) takes roughly the same position.

However, I *do* think that *both* discretionism *and* BV do a much better job of dealing with toxin-style cases, which is a surprising claim given that toxin-style cases were meant to expose the problem with discretionism and therefore for volitionism.

2.3.4: Decisions to Believe and Decisions to Do:

Here are two arguments against Hieronymi's view that do not persuade me, but may persuade others. At minimum, they raise the dialectical cost of adopting the view that discretionism divides the conceptual landscape at its natural boundaries. By that I mean something like this: It is important that Hieronymi's candidate kind of control, discretion, capture *something* of the core volitionist intuition that we are only culpable for what is up to us in the right way. If, as I have been arguing, we have discretion over some intentions, but not others, and some actions, but not others, we have every reason to think it a curious, bespoke kind of control that has no place in morally significant practices like ascribing guilt. Things would be even worse if it turned out that the *formation* of intentions (often called "decision") *were* something over which we had discretion or that the *formation* of even beliefs were too. For me it is enough that there are theories like BV that deliver much more intuitive results without incurring any obvious costs—and, as I will argue, that they do a better job of handling Hieronymi's cases than her own view. However, others may find them to be an additional reason to reject that discretion is a concept that cuts the joints cleanly.

Arguably, decisions are actions. Now, we either have discretion over our decisions or we don't. If we don't, then we lack discretion over an extremely important subset of our actions, and discretion is a still less likely candidate for what most philosophers (and most every volitionist) mean when they talk of control. If we do have discretion over our decisions, then Hieronymi's argument is in even more trouble. For, in that case, it becomes much more palatable for volitionists to simply accept that discretion is a suitable foundation for moral responsibility after all. Though it might be the case that we lack it over our intentions, our intentions (or at least the subset of them we acquire actively) are the product of something over which we do have control. As such, we will be derivatively responsible for our intentions in virtue of being responsible for the deciding that brought them into existence (or perhaps just was their coming into existence). Ultimately, then, we have

provided a means of distinguishing intentions from beliefs and other attitudes for which doxastic incompatibilists are loathe to hold us as easily and frequently responsible.

Are decisions intentional actions? For Hieronymi, they do not normally merit the label.[31] An articulation of that position might go like this: "Decision" means roughly, "coming to have an intention," and we come to have our intentions after reflecting on our reasons for acting, just as we come to have our beliefs when we reflect on our evidence. Intending is coming to have settled a answer to a question and it leads (if all goes well) to action; however, it is not itself an action.

Now, uncontroversially, there is sometimes *some* action just preceding a decision. If we need to know where to go to dinner, we might think about the relevant considerations (price, quality, distance, and so on). This we can do because we think it is worthwhile, and so over this we have discretion (for, we might think, *consideration of these factors* is made worthwhile by the aid this consideration will provide in ascertaining where to eat). This kind of conscious reflection is an action like any other. Much of it may take place in the mind (though it may also partly take place on a phone app or through conversation with one's date), but this is no worry—lots of actions are mental or partly mental, and we have discretion over those. But there might *not* be an action that is "deciding to eat *at Shezan*," though we can speak of an action described by "mulling over where to eat" or, somewhat loosely perhaps, "deciding where to eat." For Hieronymi, deciding to eat *at Shezan* is not an action, but the automatic intention acquisition at the end of some deliberation about where to eat.

Fittingly for Hieronymi, "deciding to eat at Shezan" is also not something over which we have discretion. I might choose to make up my mind about where to eat because I think doing so is worthwhile (if I don't decide, I'll miss dinner). In this instance, the action performed is one of reasoning that terminates in its aim—making *a* decision, whatever that decision ends up being. But I do not decide to eat *at Shezan* because I take deciding to eat at Shezan to be worthwhile.[32] This is why

[31] She does not *quite* say this, but it can be inferred fairly straightforwardly (2009, 177-87).
[32] Although we have already seen some reason to be suspicious that this is *always* the case in the previous discussion of intention. If Clarke is right and we sometimes have discretion-friendly reasons to intend, we will also see discretion-friendly reasons to decide crop up, especially if deciding is not an act importantly distinct from coming to have an intention.

it turns out that the relevant action, whatever it is, cannot be accurately described as "deciding to eat at Shezan," and why we need not fear we have stumbled upon a huge range of actions over which we lack discretion.

This line is an attractive one, but demanding in that is construes of practical deciding (about what to do) as very much like cognitive deciding (about what to believe). Many theorists take issue with this.[33] Mele (2003) offers the case of Joe, who is deliberating one New Year's Eve about whether to quit smoking. Joe weighs up his reasons for and against, and judges that it would be best to quit smoking. Understandably, Joe takes this as pretty bad news, since he likes smoking, but also wants to do the right thing. He announces that he thinks he should quit, and his wife then asks him if he will quit, and he replies, "Not yet! I haven't decided." Whatever Joe ends up doing, for Mele, he needs to *do something* in order to settle his future.

Prima facie at least, there seems to be no such necessary phenomenon with respect to coming to beliefs. If I am interested in the *theoretical* question of whether it is best to quit smoking, I might weigh up the reasons as Joe did, and, when I have done that for a while, I will find myself with the relevant belief. I need not do anything else, the way Joe has to. My belief that I should quit (more precisely: that it is best to quit) is immediate, his intention to quit must be mediated by some decision, and one that looks like an action.

Of course, none of this shows that Hieronymi's view is untenable, but it does show that it is costly. She must make provisions for certain kinds of intentions, establish a narrow (and potentially controversial) sense of the ability to act, and reject a widely held and intuitively gripping account of decisions.

In fact, it may be slightly more complicated. It might be that we have discretion even over belief in the sense that we can decide to *form* a belief for any reason we think makes *forming* that belief worthwhile—though it is not the case that we can *believe* for any reason makes *believing* worthwhile. I hasten to add that deciding to form a belief for reasons other than apparent truth might

[33] See Kane (1996) and McCann (1998).

not be rational; however, we have many irrational abilities. Setiya (2008) argues for a strong distinction between states (e.g. believing in utilitarianism, being tall) and what we might call "completables;"[34] namely, things that can be "finished or completed and in that sense *done* … like walking, digesting, and growing."[35] Completables are things that can be done intentionally; states are not. Believing is not completable, but it does seem that we can try to bring belief about (possibly irrationally). As such, it is the sort of thing that can be done intentionally. His argument is complex, and he is not particularly interested in Hieronymi's notion of discretion; nonetheless, he does seem to be talking about the sort of thing that might be done for any reason that makes doing so worthwhile, and therefore by discretion.

Obviously, if this is correct, no volitionist will be even slightly persuaded of the appeal of discretionsim. The bait, as it were, meant to attract volitionist intuitions was meant to be that it was a form of control we *lacked* over belief. If belief (formation) turns out to be yet another case where we sometimes do and sometimes don't have discretion, then discretion is not a form of control that is likely to appeal to many volitionists.

Hieronymi, in her response, argues somewhat persuasively that Setiya is not paying sufficient attention to a certain ambiguity in the concept of belief formation. The ghost of an outline might go like this: Her view is roughly that Setiya is confusing a kind of belief-production with belief at will. The former can be caused in the self the way a chef causes soup to come into existence—but the latter is a question of coming to have an answer to the question of whether something is true, and "you will not, by finding convincing reasons that you take only to show that believing *p* is worth doing, therein become committed to the truth of *p*."[36]

Hiernomyi herself admits that this particular disagreement with Setiya is not a merely a matter of sorting out the occasional bad inference, but is part-and-parcel to their diametrically-

[34] This is my term, not his. I am here simplifying his notion of what "corresponds to the perfective aspect" (2008,38).
[35] Ibid.
[36] (2009, 165)

opposed philosophies of action.[37] I do not intend to resolve that much larger dispute here, but merely note that anyone who strongly prefers Setiya's Thompsonian view to Hieronymi's Anscombian one will have an additional reason to abandon discretion as a useful concept.

2.3.5: Toxin Revisited:

Above, I have tried to show that discretionism, Hieronymi's initially plausible precisification of volitionism, turns out to be both highly revisionist and strongly reliant on controversial theses. Now the volitionist is in a better position to say that Hieronymi hasn't captured the sort of control in which they are interested through "discretion." But she certainly has identified *a* sort of control, and one many philosophers will think explain away volitionist intuitions about the primacy of decisions, intuitions, and actions in moral responsibility. In this section I will say something about the toxin-style cases that undergird parts of her argument. In my estimation, such cases are useful for getting a clearer picture of the foundational disagreement, especially once they are disambiguated. In what follows I argue that, paradoxically, our intuitions about toxin cases are much more strongly supportive of a broad range of volitionist views than they are of attributionist views.

When we decide, intend, act, or choose to do none of those things, we make a choice. Choice, I take it, is what BV is about. Peels prefers to speak of "uncoerced will," but since choice is a less philosophically rarified term than "will," I prefer it. A full account of volitionism would need to spell out in more detail what "choice" comes to, but its ordinary English sense gives us the desired results in all the relevant fields we have touched upon: we choose what we intend in Clarke-style cases, in ordinary intention-formation, in Velleman-style intentional action, and in ordinary intentional action. Finally, and debatably, we do not choose what we believe. That, in any case, is how I will use the term, though the reader is welcome to substitute "believe at will" over "choose to believe."

There is typically no choice in the toxin case, however. There, it looks like no choice is made, because no choice can be made. As soon as the subject realizes she has no reason to actually drink the toxin at midnight (for, recall, she will already have the no-strings-attached money she got for

[37] (2009, 274).

intending to drink it), and good reason not to, then it is too late for her to intend. She might try to think of ways to intend, e.g. self-hypnosis, promising, or other roundabout ways of getting the right intention, but as long as we rule out these indirect methods, she won't have that choice available to her.

When dealing with moral responsibility, it helps to select cases with moral stakes. Suppose that instead of offering money, the eccentric millionaire threatens to kill a large number of innocent people should the agent fail to form the requisite intention to drink the toxin. Thus, a rational agent will find herself unable to save the innocents. We can imagine such a case only with dismay! The agent, weeping and thinking desperately, eventually realizes that she cannot form the relevant intention. Surely it would be inappropriate to blame her! She was constrained by her own rationality in a way *beyond her control*.

Toxin-style cases can also be altered to force undesirable action rather than forcing undesirable inaction. We can imagine an agent presented with a cure to her child's moderately severe illness. If she now forms the intention to administer the cure when it is presented at midnight, the millionaire stipulates that many innocent people will die presently. In such a case, the agent is in a very bad situation: Whether or not she administers the cure at midnight has no effect on the lives of the innocent people, so she will have every reason to administer the cure when it arrives on her doorstep. Again, she might try to distract herself, hypnotize herself, etc.—but if we bracket those strategies, she's doomed to form an intention that will lead to the deaths of many, much to her consternation.

Presumably, such an agent is faultless. She would protect the innocent people if she could, but there is an epistemic impediment preventing her from bringing about that good end. And if this is right, what sort of counterexample does it form? Of course, Hieronymi does not offer it as a counterexample to volitionism *as such*. For her, the toxin puzzle is merely an illustration of the fact that we lack her proxy for volitionist control over intention. This, coupled with the assumption that we are directly responsible for our intentions, is what shows the falsity of volitionism. But since no volitionist would want to blame the agent above, we have yet to see a culpable intention.

Of course, many intentions are apt targets for ascriptions of moral responsibility. Ordinary, non-toxin-syle ones frequently are, and we lack discretion over those. This is sufficient reason, as we have seen, to prefer BV over discretionism, despite the latter's slightly higher level of precision. Nonetheless, it is notable that *both* seem to do better than Hieronymi's view, which does not have a mechanism for denying culpability in these cases, at least as far as what she says goes.

There is a charitable way to imagine how she might respond. Hieronymi, like most attributivists, is happy to ascribe praiseworthiness and blameworthiness without choice. So, in the case of the mother with the sick child, she might say the mother is *doubly directly praiseworthy* for what happened, because her intention-formation was attributable to her maternal care, and her consternation at the death of innocents was reveals an appropriate regard for the value of their lives.[38]

Volitionists of every stripe would surely allow that the mother is indirectly praiseworthy for her virtuous inclinations towards care and consternation, so long as she had a hand in bringing those about through past choices. But it gives a more intuitive response to the matter of culpability for the deaths: The mother is excused, because she could not choose to save the innocents due to an epistemic constraint beyond her control; namely, her knowledge about what she *would* do with the cure when presented with it later. To be clear, I do not think this is decisive advantage for volitionists, but it does seem that an excuse is more appropriate than praise. The mother in the case is a good woman; she has done nothing wrong; what bad things she has brought about are excused, because her power to choose was undercut by self-knowledge beyond her control.

Friends of attributivism might look for a better case in which (1) there are moral stakes, (2) in which the agent is clearly morally responsible, and (3) in which the volitionist cannot appeal to choice. If the agent there is directly blameworthy, the problem for the volitionist reemerges, for direct culpability and choice have again come apart.

Hieronymi does offer another case:

---

[38] Recall that in the case, the mother did *not* choose a moderate improvement in her child's health over the lives of many innocents. Rather, her regard for her child caused the deaths of the innocents through a toxin-style mechanism.

> Perhaps I am unhappy that you do not intend to marry me—and perhaps I care less about the social and legal arrangement than about your state of mind, the fact that you have no such intention. And perhaps you would be willing to adopt a state of mind that pleases me—perhaps you would be willing to house a marriage commitment, so to speak—so long as you do not need to actually engage in the matrimonial and enter the legal condition (2008: 268).

She does not elaborate further regarding culpability or fill in every detail. Either choosing to marry is something you can do independently of toxin-style considerations, or it is not. Suppose first, that it is. In such a case, choosing to marry (or not to) is up to you. It is true that you may still lack *discretion* (on Hieronymi's technical definition of the term) over the *intention*, in that you can only intend for reasons that make *doing* so worthwhile, but not (just) for reasons you think make *intending* worthwhile. But it is not at all clear why this is a problem for volitionists. Suppose, instead, that you are unable to choose to marry for the toxin-style reasons. To fill in the details of the case: Hieronymi, on this reading, holds this to be a toxin-style situation (a "toxic relationship," perhaps). So, before you can give the question of whether to marry a think, I tell you that I don't actually care whether we go through with it—just that you form the intention.

Here's a very congenial way of framing your response: "If you wanted to get married, I would. But you don't seem to, so I can hardly be expected to intend to." It would be manifestly unreasonable of me to reject this response and resent you. Suppose, instead, you reply something to this effect: "You know I want to make you happy, but I just don't think marriage is a good idea." But notice that this reply is compatible with the first. *Neither* fully implies what Hieronymi stipulates; namely, that you would not be willing to get married even if it *were* important to me that we actually marry. As it stands, what I am being blamed for is not something voluntary, but my *willingness*. As Hyman (2015: 80) points out, in war there are willing conscripts and reluctant volunteers. The question is partly whether the conscription provides an excuse (at least from *direct* blameworthiness) and the volunteering that grounds praiseworthiness or blameworthiness.

Willingness is not quite what we should be getting at, however, for it is an attitude (of some sort), *not* a failed intention. The volitionist was never going to allow direct moral responsibility for attitudes; Hieronymi wanted *intentions* to serve as the trojan horse through which direct responsibilities for attitudes would be smuggled in. She has not, however, succeeded in this particular

case. Though you may be willing to marry me in the sense that you might be perfectly happy to, you might not intend to. How could it be that I could blame you for not so intending (for I do not actually care if we marry), except insofar as I am using your intention as a marker of your general willingness? And, since general willingness frequently comes apart from our actual intentions, she has not succeeded in offering a case of a blameworthy intending (or failing to intend) that the volitionist cannot account for. I think, for this, reason, that it is better to move away from her example.

Fortunately for the volitionist, the revised toxin case offered above already showed a blameless agent. Perhaps another can be constructed to be maximally sympathetic to Hieronymi. This, at any rate, is what I have tried to construct.

Suppose Francis owes his business partner a million dollars. Since he is selfish and duplicitous, he has no strong desire to pay, though he does somewhat desire to do so, for he is not wholly wicked. One day, he's confronted with a toxin-style proposition. He's told he'll be given a million dollars (no strings attached) if he can intend to use it to pay back his partner. Francis knows that if he gets his hands on the million he can immediately place it in a secret offshore account to use later for buying yachts. In fact, he would much rather do that than pay back his partner, though he would much rather pay back his partner than receive nothing at all.

Toxin-style cases are strange, and we don't encounter them very often. But I think there are at least two ways they can go, psychologically speaking. One is culpable, but, crucially, it is so in virtue of smuggling in choice, and, as such, is no threat to BV. Here are two ways Francis' thought process might go:

1) "Ah, this billionaire has some lie-detector he wants to hook me up to, and some strange conditions about giving me a million dollars. Gosh, that's a lot of money. I wonder what I should do if I get it? I could pay back my partner. I do feel a little bad about having stolen from her. But that's life! She shouldn't have been so naïve. I'm totally going to keep it, if I get it. Now, what were these conditions again? … Ah, alas…"

2) "Let's see, so I only get the money from this idiot if I intend to repay her, but he's not going to make me sign an enforceable contract. So I wouldn't have to repay her! And, if I know me, there's no way I'd give that fool my money! But wait, that means… oh no!"

These are not exhaustive. But they do offer the volitionist a way of saying why it feels like Francis might be to directly to blame for something. In the first case, Francis made a choice. He decided something—conditional on his getting the money, to be sure. But we form conditional intentions all the time, so this is no trouble. Now, most volitionists do not think someone's making this kind of decision is *sufficient* for her being moral responsibility, though it's a good place to start. Perhaps indeterminism must be added, or epistemic conditions, or whatever. But those debates are neither here nor there. What may well be the case is that it initially *feels like* Francis made a choice—that he decided something. If that is indeed how the details are filled in, then the volitionist has a wide open avenue for ascribing culpability. If, conversely, we specify that no choice is ever made, then the case loses its intuitive punch. Admittedly, the attributivist might point out that the reason that no choice was possible lies with Francis' character. This is correct, as far as it goes. His being unable to pay back his partner *is* reflective of his character, for what is blocking his capacity for forming an intention is his rapacious character. This, however, has all the hallmarks of the debate prior to Hieronymi's entry. We tend to think that nasty people are blameworthy for something, but the volitionist always has *derivative* culpability to fall back on. Perhaps Francis is deeply blameworthy because Francis made himself into the kind of appetitive wretch he is, and that explains why we want to blame him. What she has not done is offered a successful *new* rebuttal.

Nevertheless, it may seem to some that Hieronymi has provided a case of clear *direct* culpability, absent any choice. But any account of moral responsibility worth its salt will make provisions for some epistemic conditions on its exercise. If I am *completely* certain that I cannot make

a half-court shot, then, if I accidentally do, I surely do not do so intentionally.[39] It thus makes it so that I lack discretion over doing it.

If this is right, then the same conditions should be accessible here. In Francis' case, he thinks he is unable to form the relevant intention, in this case because he believes he wouldn't pay back his partner. But, nevertheless, it remains true that what is blocking his intention is a belief—namely about what he would do once he had the money.

But if it is reasonable to build this kind of epistemic condition into discretion, why not for other candidate forms of control? This is to say, Francis has the same kind of excuse that I do when I fail to intend to make the half-court shot. I cannot decide to make the shot because I believe that I am not that great a basketball player, try as he might. Francis cannot decide to form the intention, try as he might,[40] because he believes repaying his partner isn't the sort of thing he can do, given who he is.[41] One might reply that what renders Francis unable to get the money is really his wicked character. But his wicked character alone is not what renders him incapable. He might get his self-assessment wrong: perhaps he is unaware of the rapacity awoken in his soul when he touches large sums of money, and takes himself to be the kind of fellow who repays debts. In such a case he *can* get the money. Or conversely some agent, Justine, might be very generous and dutiful, but have an unfairly low opinion of herself. That agent, put in Francis' shoes, would fail to get the money, *not* because she is wicked, but because she is mistaken about her own character. For all that, she will fail to get the money, *even though* she is capable of forming the relevant intention, *because of a belief about herself*.[42] The fact that Francis is—on some level—capable of repaying his partner is no more relevant

---

[39] Some might reject this view for reasons related to the earlier discussion of Davidson's carbon-copies. But an extreme Davidsonian view (more extreme than his own) is a double-edged sword, for it would entail that Hieronymi's account of intentions falls apart, as explained above. It is a problem for her in either case.

[40] Remember, Francis does want to take the deal, so to speak. He prefers paying back his partner to getting nothing, and he sees those are the two options.

[41] I am *not* here claiming an exact parallel. The belief blocking me from φ-ing when I attempt a half-court shot is a belief about the impossibility of my φ-ing. The belief blocking Francis from intending to φ is a belief about what he would do, which in turn renders him incapable of φ-ing. Rather, if the volitionist is allowed to replace discretion with a different kind of control, it would be unfair of Hieronymi to deny her access to an epistemic condition eschewing responsibility for those who think themselves unable to bring about the relevant state of affairs. The epistemic condition counts.

[42] Recall that it is certainly the case that Justine might form the intention anyway, depending on the order and way she thinks about the case. But then she has made a praiseworthy decision, and volitionists will have no problem praising her. In the case I am considering, she is incapable of forming the intention because she first

to responsibility here than the fact that I might be—on some level—capable of making the basket. Very plausibly, neither can intend, and neither is *directly* responsible for that omission.

I want to be clear about what I take myself to have argued. For all I have said, attributivism is correct. I have not argued against it. Rather, I have argued that Hieronymi's trojan horse strategy has failed. Any volitionist who was unimpressed before will remain unimpressed, for Hieronymi has failed to deliver real counterexamples to the volitionist view. In neither her own cases, nor those I constructed on her behalf, do we find any cases of directly responsible agents lacking choice. She may, however, have helped make it more precise. What of "discretion," as Hieronymi construes it (Recall: An agent has discretion of φ just in case she can φ for any reasons she thinks makes φ-ing worthwhile)? Well, it is perhaps true that there are limits on our discretion over our intentions. However, the only cases of blocked intentions were cases the volitionist is not inclined to think admissible in the first place. It might be better to stick to the language of choice, and to freely admit that *some* intentions (or failures to form intentions) are beyond our powers of choice. They can be blocked by our believing that we cannot perform the action to which the intention was directed. This, however, is entirely unsurprising, and is acknowledged by any moderately sophisticated theory of action since Anscombe.

2.3.6: Why Toxin cases matter:

In some sense, most of the previous discussion of toxin-style cases does not matter very much. The volitionist has every reason to prefer BV over discretionism because discretionism has trouble dealing with *normal*, non-toxin-style intentions. We cannot form intentions for any reason we think makes doing so worthwhile, at least normally. Why worry about such a small set of cases?

Hieronymi might have a dialectical reason to favor a certain class of toxin-style cases in which no decisions are made, nor any actions undertaken. Since, arguably, deciding is itself an action, were Hieronymi to stray from this narrow set, she risks giving the volitionist back all the moral

considers the conditions necessary to get the money, rather than making a decision about what she will do with it if she does get it.

responsibility she wants through the back door: The volitionist might simply accept that we are not responsible for our intentions directly, but then insist that we *are* directly responsible for our decisions, and thereby derivatively responsible for the intentions those decisions create. I unsure whether this strategy is persuasive, but it is the sort of thing a reasonable person might think. Even if decisions fail to clear the bar of discretion, there is clearly *some* kind of volitionist-friendly control at work in ordinary intention-formation/decision. Volitionism is the view that responsibility comes down to what we *choose* or *do* or *will*, etc., depending on how they wish to work out the details. It is only on a very specific construal, discretionism, that volitionists face a problem. As such, she has every reason to examine cases in which we are much less likely to have this broader kind of control. However, as I hope to have shown, volitionists face no problem here.

Hieronymi's argument was designed to convince us that we are morally responsible for our beliefs and attitudes, because they are relevantly up to us. As such, it is argumentatively suspect to deny Francis an excuse on the basis that the belief (and underlying attitude, for that matter) are somehow within his control. Presumably, we do not want to blame Francis for his accurate self-knowledge. It is a good thing to know oneself. Whether Francis is directly blameworthy for his character is, of course, the central question of the whole debate. I have not argued that Hieronymi has given an incoherent view or an obviously flawed one. Rather, I have sought to show that in toxin-style cases agents usually lack *both* discretion and more obvious sorts of volitionist-friendly control, while in ordinary cases of intending, agents can decide or choose as they see fit, and thus might be responsible. So, while toxin cases do show that we lack discretion over our intentions, they do not show that we lack other volitionist-friendly sorts of control. Nonetheless, she has provided a useful tool for sorting views into the respective volitionist and attributivist camps. Those inclined to think Francis directly blameworthy in all cases might find attributivism appealing, those not so inclined probably won't. Her argument, while clever, does not shift the debate.

Rather, if anything, attributivism actually looks slightly worse-off, while volitionism is helpfully precisified. The latter is obvious—ruling out her form of volitionism makes it easier to choose an alternative. Her work is in that regard very helpful to everyone. As regards the former

claim: I have argued that attributivists have a less intuitively-appealing account ready to hand for dealing with good agents bringing about appalling consequences without choosing to do so. In the case of the mother and the medication, it seems more natural to say she faultlessly caused something bad to happen through her intention than to say she is praiseworthy. Of course, a sophisticated form of attributivism may not be committed to saying as much, and I will leave that work to them. For my purposes, the primary takeaway is that, while discretionism is likely false, volitionism is none the worse off.

2.4: The Second Unanswered Challenge:

I now turn to the second of our three responses on Alston's argument detailed in the introduction. While this argument, taken from John Hyman's influential *Action, Knowledge, & Will*, is not explicitly levelled at either Alston's argument or the "volitionist" position, it is broadly against the spirit of both, as a little work will show. I argue that Hyman's core argument, while original, nuanced, and clever, fails. Hyman seeks to totally reorient our understanding of free will and moral responsibility around the concept of the involuntary, where "involuntary" entails suffering from compulsion or ignorance. Very roughly, we are morally responsible for some state or act if and only if that state or act is not enacted or suffered out of duress or ignorance. Hyman does not commit himself to a particular view on the voluntariness of belief, but, if I am right about what his view entails, then he is opposed to Alston's defense of the bad believer. If he is right, it is possible to be directly morally responsible for a belief that is not chosen. Indeed, on his picture, a belief's being chosen has little or no bearing on our being morally responsible for it, since he also thinks that a belief can held involuntarily even if chosen.

I begin by laying out Hyman's view and his main defense of it. Then I apply his view to the case of belief. Doing so is not much of a stretch, but does require a little extrapolation and elaboration. Finally, I will show how his argument fails by falling prey to the same onerous requirements it imposes on other views.

2.4.1: Hyman's view of Voluntariness:

Hyman begins with the observation that, *pace* Ryle, Wittgenstein, Davidson, Kenny, Bernard Williams, and later Anscombe, voluntariness applies to more than just action. He approvingly quotes a short passage from Anscombe's early phase, in *Intention*: "Things may be voluntary which are not one's own doing at all, but which happen to one's delight so that one consents and does not protest or take steps against them: as when someone on the bank pushes a punt out into the river." Later Anscombe (seemingly) changed her mind, assuming voluntariness can be defined in terms of knowledge, intention, and self-control. She was mistaken to do so, because "voluntariness is at root an ethical concept, unlike knowledge, intention, and self-control" (76). Its basic function is to "inform the appraisal of individual conduct and in particular of the concept of innocence and guilt" (76).

For Hyman, the area of philosophy that deals with moral responsibility ought to be fundamentally concerned with figuring out what is voluntary. This is assumed. The problem up until now, in his mind, has been over-attention to e.g. intentional action. Attending to the ways in which passive states can be voluntary reveal as much. But once we give up the chase for a positive account of the voluntary based in knowledge, intention, and self-control, we can come to see what the voluntary really is:

> What is voluntariness? Roughly, a certain thing is done voluntarily if, and only if, it is *not* done out of ignorance or compulsion. This is not the same as saying that it *is* done knowingly and freely. Regarding knowledge, if a man drives the wrong way down a one-way street, and it transpires that although he believed the street was one-way, he did not *know* that it was… he cannot be excused on the grounds that he did not do it voluntarily… But what does ignorance have to do with compulsion? Why do we have a concept that yokes together these disparate ideas? The answer, I suggest, is that both are normally exculpations… I suggest that concept of voluntariness is formed by negation, in fact by double negation, by excluding factors that exculpate, in other words, factors that exclude guilt… The factors that negate voluntariness are ignorance and compulsion, regardless of whether the thing whose voluntariness is in question is an act, omission, event, relation, condition, and so on. Aristotle states this explicitly with respect to action and emotion. But the idea that ignorance and compulsion are exculpations was already ancient in the classical age, and it does not seem to have originated in any single ethical tradition. Indeed, it is hard to see how the ideas of guilt and innocence, as opposed to the simple idea of causing harm, could exist without it (77).

Hyman's view is clearly enough stated, but its distinctiveness is not very salient until he elaborates on the relationship between voluntariness and choice. This relationship is crucial for our purposes. Recall Alston's view (put here colloquially) that we lack direct moral responsibility for our beliefs because

we believe without choosing to believe. Since we cannot do otherwise, we have an excuse if things go wrong. Now, Hyman readily admits that voluntariness is closely connected with choice in some way ("[It] is natural to think that a person does something voluntarily if, and only if, he chooses to do it, or at least could choose not to do it" (81).). But examination of duress presents a difficulty.

Lord Widgery, a former Lord Chief Justice, defines the English legal position on duress as follows:

> It is clearly established that duress provides a defence in all offences including perjury (except possibly treason or murder of a principal[43]) if the will of the accused has been overborne by threats of death or serious personal injury so that the commission of the alleged offence was no longer the voluntary act of the accused (R v Hudson, R v Taylor [1971] 2 QB 202, 206, quoted in Hyman 82).

On one conception of duress, choice and voluntariness start to come apart. On one way of understanding the word "choice," the presence of some kind of threat undercuts voluntariness, and therefore guilt. But, of course, in another sense, choice is still (sometimes) present at the same time as duress. As Hyman points out, Lord Widgery does not mean, in the quotation above, that the accused's will is "overbourne" in the sense that the accused is unable to act otherwise. The accused may die if he acts against the will of his threateners, but people sometimes choose to die. As Hart and Honoré note, "the person threatened is literally able to choose whether to act as instructed or suffer the threatened harm, so that a choice exists" (82). This is perfectly consistent with recognizing that duress does not *always* leave some kind of choice intact. It might be that *some* kinds of duress *do* literally render agents unable to do otherwise, as in cases of severe torture. But, bracketing those, we might take what Hyman calls the "Nicomachean" view, according to which x is voluntary if and only if x is chosen.

> Nicomachean View: Duress completely vitiates voluntariness only when it renders agents literally unable to choose (91). E.g., if an agent's tortured to the point where he cannot choose not to spill his secret, he doesn't spill his secret voluntarily.

[43] This exception will turn out to be important, so I flag it here.

Alternatively, we might stick with Lord Widgery, as Hyman would have us do. This, he says, is the view consistent with Aristotle's *Eudemian Ethics* and echoed later in Aquinas.

> Eudemian View: Duress can sometimes completely vitiate voluntariness, even when the agent in question is "literally able to choose" to do otherwise (2015, 81-90). E.g., if an agent's threatened with serious bodily harm unless he hands over his wallet, he doesn't hand over his wallet voluntarily.

One more clarification is needed before Hyman's view comes into sharp focus. We have so far seen how the voluntary comes apart from choice. But there are multiple ways of understanding this kind of freedom from duress and ignorance. One way concerns what Hyman calls "willingness." Suppose you threaten me with a bludgeoning if I refuse to do a little dance. As it happens, unbeknownst to you, I have no aversion whatsoever to dancing right now; in fact, I think it's a great idea. In this case I'm still dancing under duress in Hyman's view, because of the circumstances in which you have placed me. Nonetheless, I am, as it were, a willing dancer; just not a voluntary one. Suppose, conversely, that you ask me to dance, without threatening me. Now suppose that I hate dancing, that I am feeling nauseous, and that I loathe you. It might *still* be the case that I decide to dance with you, despite all this, because, say, I take party etiquette extremely seriously. In this case I am dancing *voluntarily* (no compulsion or ignorance in sight!), but I am doing so *reluctantly*. A willing conscript is still a conscript in Hyman's legal-theory-informed eyes; a reluctant volunteer is still a volunteer.

Here is a summary of terms as Hyman uses them:

1) Some act or state is *freely chosen* only if it is chosen and one could have refrained from choosing it.
2) Some act or state is *voluntary* only if it is not the result of ignorance or compulsion.
3) Some act or state is *willing* (or *done willingly*) only if it is not done reluctantly.

I am not interested in whether these are correct ways to use English words. I am interested in whether Hyman's view is correct, and in the implications it has for responsible belief. Before we get to that, we need to look at Hyman's reasons for favoring the "Eudemian view." The first is implicit: It seems

to be the favored view of the Western legal tradition, and this is at least some reason to take it seriously. Lawyers and judges have been thinking about justice for a long time, and deserve to be taken seriously. I however, lack the legal and historical knowledge to comment on this argument intelligently, so I will ignore it.

The main argument is in the form of a reductio of the so-called Nicomachean position. Any crime of force (e.g. extortion, unjust conscription) can serve to illustrate it; Hyman chooses rape, perhaps because of the force with which that heinous crime grips us:

> And it is clear that coercion and voluntariness are opposed, where the patient of an act is concerned, because whether an act is voluntary on the part of the patient depends on whether it is done with her consent, and a sufficiently grave threat vitiates consent. In some cases it is a matter of judgment how grave the threat must be. But if a rapist threatens a woman with violence and she makes a conscious decision to submit—as she may, if the circumstances allow her to weight the risks involved—the idea that she must therefore have consented to the act and submitted to it voluntarily is repugnant and absurd, regardless of whether she was "literally able to choose" whether to submit or risk being killed… [The] advantage of the *Eudemian* doctrine is [therefore made] obvious. For the *Nicomachean* doctrine requires us to say that if terror and panic made the victim literally unable to resist—in other words, if her submission was "made inescapable by forces within herself which she was unable to overcome"—then she did not consent to the act or submit voluntarily, whereas if she made a conscious decision to submit because she judged resistance to be too dangerous, than she did consent to the act, and did submit voluntarily… So if the *Nicomachean* theory were right, it would be a defence against the charge of rape that the victim was "literally able to choose whether to act as instructed or suffer the threatened harm" (89).

It is plainly absurd to argue, as a defense against the charge of rape, that since the victim chose not to fight that she acceded voluntarily. Since this is what the Nicomachean view holds, we ought, he says, jettison it.

2.4.1: Application to Belief, and the Ineluctable Revenge of Choice:

As I mentioned above, Hyman does not explicitly extend the language of voluntariness to belief. He does say, repeatedly, that it extends beyond action. Examples are numerous. Recall: "The factors that negate voluntariness are ignorance and compulsion, regardless of whether the thing whose voluntariness is in question is an act, omission, event, relation, condition, and so on" (77). And again: Voluntariness "can be attributed to several things apart from action: conditions such as exile and unemployment; attitudes, such as forgiveness; being acted on by another agent—for example, being

tickled, kissed or killed; inaction or quiescence; failures to act, such as neglect; relations, such as occupancy and possession; and events that can occur by choice, such as when one falls asleep or dies (88). These are all passive. Whether they are voluntary will depend on the specifics of the case, but I think it safe to say that most (perhaps all) involve some kind of choice. When someone voluntarily dies, it is not the case that she intends anything when she is dead; the choice occurred earlier on. But there was a choice. Hyman does not, insofar as I can tell, come down very clearly on whether he thinks some kind of choice is a *necessary* component of voluntariness, though, as we have seen above, he is very clear that it is not a sufficient one.

He does get close at one point. "If someone does something voluntarily, it does not follow that it is possible for him to avoid doing it" (98). He goes on to cite Locke's famous example of the man who, not knowing that he is locked in a room, decides to stay in the room. "The man in Locke's example chooses to stay in the room, and he does so without being coerced; but in fact, although it is a fact that he is not aware of, he does not have a choice whether to stay or go. The choice he does have is whether to *try* to go, and whether he stays voluntarily depends on which choice he makes, since he will be compelled to stay if, and only if, he tries to leave" (99). This position is compatible, insofar as I can tell, with a standard volitionist approach. Most volitionists would be perfectly happy to say that Locke's man stays in the room voluntarily, and that he is (all other things equal) morally responsible for doing so. But they would say he is responsible for doing so *in virtue of having decided to stay*. That is, he is indirectly responsible for being in the room in exactly the way Hyman describes. This, anyway, is one way of understanding what Hyman is getting at.[44]

My sense is that most people believe they decide what they believe in a way similar to the way they decide what to do. I think they are wrong about this, but they think it all the same. But if people think they can control what they believe, then they are analogous to Locke's prisoner. Just as he could decide to leave the room, so they can decide to change their minds. Would it not follow that they could be responsible for not choosing to change their minds? More carefully: would it not be true

[44] See Frankfurt (1969), Widerker (1995), and Mele & Robb (1998) for canonical discussions of the principle of alternative possibilities.

that they are obliged to try to change their minds in certain circumstances, and would be culpable for not doing so? I think the answer must be "yes." But note that it does not follow that they are directly morally responsible for believing as they do, merely that they are (or might be, other factors permitting) directly responsible for not trying to change their minds, and, perhaps, indirectly responsible for believing as they do. By analogy, imagine that an innocent person lies imprisoned across the hall. You have a key to the room, and you lose nothing by freeing them. Now, as it happens, your key is broken, but you are unaware of this. If you persist in doing nothing, we have grounds to think rather less of you. Once you realize your key is broken, inaction becomes permissible again. If this is correct, it motivates an interesting, if rather small, amendment to Alston's dictum. It is not the inability to choose otherwise that excuses bad beliefs, but the realization of that inability.

In any case, recall that Hyman's account of moral responsibility is negative. We *act* voluntarily only if we act free from ignorance or compulsion. We also *abstain from acting* voluntarily only if we do so free from ignorance or compulsion. What does it mean to *believe* free from ignorance or compulsion? Ordinarily, it seems that belief is impregnable from compulsion. If you have good evidence that it is a Tuesday, no matter how I might threaten you, you will still believe it is Tuesday (though you may lie about it, say). As for the ignorance excuse; well, if ignorance sometimes excuses bad action, does it excuse itself?

Hyman does mean his negative conditions to be generalizable as we have seen. Recall: "The factors that negate voluntariness are ignorance and compulsion, regardless of whether the thing whose voluntariness is in question is an act, omission, event, relation, condition, and so on. (77)" Having the wrong belief, or wrongly lacking a belief more-or-less just *is* ignorance. Is ignorance itself excusable only if it is the result of more ignorance? Presumably not—or a regress would arise.

But it is not the case that Hyman thinks that ignorance is always and everywhere excused. Negligence, he stresses, is culpable (78, 80), though he gives no account of it. Volitionists have an account of negligence, recall, that fits very nicely with their general view: We behave negligently when we have a choice between taking care and do not choose to take care.

Hyman's account will likely not be as nicely generalizable. He cannot explain why some cases of ignorance are excused and others are not by recoursing to the view that "ignorance excuses." His view is less readily generalizable.

There may perhaps be a kind of general strategy available to Hyman. Here's an attempt: Let us say that someone believes voluntarily just in case she does not to do so involuntarily, where "involuntary" is the disjunction of "believes under duress" and "believes under ignorance." What might this look like, and can we succeed in sidestepping, as Hyman wishes to, the question of choice?

Well, it is certainly plausible that we can fall prey to certain kinds of epistemic duress. Extreme stress, torture, betrayal, cruelty, and so on can all have doxastic effects. In the spirit of Hyman's work, we might say that, given sufficient mental distress, certain false beliefs are involuntary. Classic examples might include the belief that one's spouse is not having an affair, or that one is not evil, despite manifest evidence to the contrary. But a sort of intuition check might show that it is not the case that we actually change our beliefs in any clearly directional way in response to duress. In these cases it looks like the agent in question may be deciding not to acknowledge a belief, to lie to herself, or to refrain from thinking about a question. But these phenomena, even if they do lead, in the end, to an actual change in belief, are, yet again, cases of *tracing* a belief back to a previous choice (e.g. not to think too much about why one's husband is taking so many business trips). A focus on what I have been calling "external," i.e. those elements of moral responsibility, like threats, that occur outside the head, keep finding their way back into the head. We cannot speak sensibly about duress with respect to belief without appealing to the psychological anguish involved, and an examination of that anguish reveals the ways in which the mind *takes certain actions* to protect itself.

What of the ignorance excuse? All bad beliefs, or at least almost all, are instances of some sort of ignorance. But even the volitionist, already squeamish about assigning blame, does not excuse ignorance just in virtue of ignorance's being ignorance. Rather, she excuses it from direct moral responsibility, roughly speaking, because ignorance cannot be helped. When it can be helped, and sometimes it can be helped, there is no excuse. But in those cases, the "help" involved necessitates the

agent choosing to go learn more. Hyman could borrow a volitionist account of culpable ignorance—or avail himself of other views on the market. But, as a consequence, his account of voluntariness does not generalize to states like "believing" in a clear way.

Parsimony is not the only virtue, but it is a virtue. It seems that Hyman is committed to saying that acts and omissions are voluntary lest they count as ignorant or compelled; but it is not at all clear what we say about doxastic states on this view. Is ignorance compelled? On volitionism, doxastic states are seldom, if ever, chosen. Therefore they are not the sort of thing for which we are *directly* responsible. Hyman is welcome to say something similar, but why commit himself to identifying voluntariness with choice here when he thinks that identity relation does not exist elsewhere?

Hyman could also help himself to attributionist accounts, according to which we are responsible for doxastic states just insofar as they are attributable to us in the right way. But what does this have to do with ignorance or compulsion? Attributionists have a positive account, and they justify it with positive argument. I am responsible for what I think because my thoughts disclose *me* in the right way. Should they fail to do so, I am not responsible. Exceptions to responsibility follow from the positive account, not the other way around, as in Hyman's account. Considerations of parsimony and consistency then support traditional "positive" views, or at least they do until Hyman works out his negative view in more detail.

Of course, none of these worries about Hyman's account address his primary argument against Volitionism; namely, his *reductio* concerning cases of force. Even if his own view were entirely wrong, his critique must be answered separately. It is there I now turn.

2.4.2: The Revenge of Choice, Once More:

Let us take a moment to recall Hyman's terminology:

1) Some act or state is *freely chosen* only if it is chosen and one could have refrained from choosing it.
2) Some act or state is *voluntary* only if it is not the result of ignorance or compulsion.
3) Some act or state is *willing* (or *done willingly*) only if it is not done reluctantly.

Hyman's contention is that free choice isn't enough for sexual consent. If freely choosing to have sex over other options were enough to make it voluntary, then someone who can resist at great personal risk has sex voluntarily when she chooses not to resist. Only a vile legal system would accept the following defense: "She could've chosen to resist. I would've beaten her, but she probably would've escaped. Since she'd a choice, she acted voluntarily. So the sex was consensual." As Hyman notes, the victim's judgment that resistance would be too dangerous doesn't imply consent (87). Rather, it implies that while there was a sense in which she chose her own course, that choice is insufficient for voluntariness.

If this is right, we have a basis for rejecting what Hyman calls the Nicomachean View, the volitionist view that free choice is sufficient for voluntariness. The Nichomachean view isn't as badly off as it seems, I will argue. However, we should consider Hyman's alternative inspired by Aristotle's *Eudemian Ethics*. The Eudemian view shifts the explanatory burden in such cases onto duress. It is therefore to duress that we now turn.

2.4.3: Crimes of Force or Extortion:

Sexual violence is a non-ideal test case for a few reasons. First, it is highly emotionally charged. That is not to trivialize other crimes of compulsion, including the one on which I will focus, unjust conscription. Many have faced brutal consequences should they refuse to participate in unjust wars, and their predicaments are nothing to be sneered at. However, mercifully, conscription is relatively rare today, while rape remains a reality for women in every stratum of every society.

A second reason not to focus on rape concerns how inexplicit many sexual agreements (or non-agreements) can be. Most people do not spell out sexual contracts on paper; they simply have sex. As such, the law must be very careful to deal with many kinds of manipulation and deception. To take a few examples:

1) Sometimes men pretend to be people other than who they are in order to obtain sexual favors from women.

2) Sometimes men pretend to have qualities they do not have in order to obtain sexual favors from women.

3) Sometimes men dangle implicit reward incentives to needy women.

4) Sometimes men level implicit threats (for example, non-verbally, through tone).

The law must do its best to ensure bad conduct is illegalized, but that innocent miscommunication is not punished overly harshly. None of this is to say that a response cannot be made in response to Hyman in the same terms in which he lays out the case; namely, sexual ones. But since the general point can be illustrated through a number of examples, I will help myself to a different one.

In what follows, I examine cases in which agents are under genuine pressure, but pressure insufficient to justify their acquiescence. We can imagine cases in which agents are pressured with various incentives in a morally detestable way, but not such strong incentives that they are blameless for acquiescence. The situation is, in principle, generalizable. It might apply to financial extortion, say, as well as extortion of the sexual or military kind. I will proceed with a military example in which a victim is placed under duress in an attempt to motivate him to fight for an unjust cause.

As a note: the matter is generalizable enough that there need be no threatening agent at all. Duress may stem from agential or non-agential threats. To borrow Aristotle's example, a ship captain throwing his cargo overboard to save his ship from a storm acts under duress not because another agent is threatening him, but because the world has rendered doing so the only acceptable option. Likewise, a man desperate for money may sell unjustly violent services because the starvation of his children is an unacceptable option. When the threat in question is agential, it may be violent or non-violent. If some cretin threatens to ruin a man's reputation if he refuses to fight for an evil cause, he puts the victim under duress, but doesn't do so by threatening any kind of physical violence. What matters here is whether the threat posed by certain harms, be they violent and agential or not, is enough to put the agent threatened in a state whereby his subsequent submission counts as non-voluntary.

Indeed, building the absence of duress into the definition of voluntariness helps us avoid the alleged problems with the Nichomachean view. It would be controversial to claim that a man threatened with extreme violence fights voluntarily, even if he deliberately chooses fighting over other possibilities. When one *φ*-s, simply having other options to *φ*-ing doesn't make *φ*-ing voluntary. One must have other *acceptable* options.

Crucially, however, determining whether an agent has acceptable alternatives isn't merely a matter of determining whether a threat is sufficiently unpleasant or harmful or subjectively disliked. According to any reasonable view, duress has limits, and examining these exposes the cracks in Hyman's account.

The important thing to notice is that whether one is under duress isn't merely a function of the harm threatened; it also has to do with the *stakes* involved. When the threat is great and the stakes low, giving into the threat looks involuntary. If I threaten to punch you very hard in the face if you refuse to do a demeaning little dance, then your subsequent dancing would be involuntary, even though you can refrain. Being punched in the face is awful, and while doing a demeaning little dance isn't fun, you have no acceptable alternative. On the other hand, if we hold the threat fixed but raise the stakes, voluntariness returns. If I threaten to punch you very hard in the face if you refuse to murder your whole family, and you do murder your family to avoid being hit, no sane court of law or public opinion would accept duress as a legitimate excuse. Hyman explicitly says as much: "[W]hether we accept a person's claim that he was obliged to give way to a serious threat depends both on the severity of the threat and on the value of the interest sacrificed by giving way" (98). In these cases, the "interest sacrificed" by acceding makes all the difference. Lord Widgery, recall, stops short of saying that duress *always* vitiates voluntariness even when *death* is on the line. Treason and murder might *not* be excusable, even under a threat as grave as murder. Presumably, this is because treason and murder are extremely high-stakes crimes.

If some acts count as voluntary even under the threat of murder, it's plausible that serving in a military effort, even in a non-violent capacity will count as voluntary under some lesser threat. Therefore, it's false that a victim's enlisting in an unjust war under threat counts in *all* cases as

involuntary. On the other hand, it's easier to see that the victims in such contexts still count as—well—victims, albeit with some level of voluntariness stirred in.

A case demonstrating this isn't easily constructed for this reason: When we want an action to count as voluntary, we need the stakes to be high relative to the threat. Conscription cases are not particularly difficult to construct in this way.

Possibly, the stakes in conscription are always high. But even if military effort isn't *always, everywhere* so important, there are *some* cases in which enlisting is a big deal. It is, at least sometimes, very high-stakes. My strategy is to make the stakes so high that giving into a *comparatively* low-grade threat counts as voluntary:

> "Voluntary" Conscription[45]: Rodong opposes the viciously repressive regime in his nation. In particular, he knows that the state secret police brings about unbearable hardship for his countrymen. He further knows that his powerful uncle, Hyunbin, works in the Ministry of State Security and wants Rodong to join up after university abroad. Rodong has long looked forward to spending some time abroad; in fact, it is his life's dream. On the other hand, he wants nothing to do with his uncle's organization. He knows that Hyunbin won't do anything to seriously endanger his life or liberty, but when Hyunbin threatens to take away his chance at university abroad should Rodong refuse to join the secret police, Rodong buckles, and agrees to sign his enlistment papers.

My contention is that in this case Rodong acts "voluntarily," as Hyman uses the term. For Hyman means by "voluntary" the class of actions for which an agent is morally responsible[46] and Rodong is morally responsible for joining up. His choice manifests a lack of due regard for the oppressed citizens of his nation. So, Rodong acts voluntarily. Had the threat been sufficiently strong

---

[45] I am using "voluntary" exactly as Hyman does. I do not think, nor does Hyman in any personally meaningful sense, nor should anyone, that being raped is ever voluntary in any sense that implies the victim's consent. Rather, it is "voluntary" in this case, if "voluntary" means what Hyman thinks it does.

[46] The language Hyman uses to describe voluntariness makes frequent use of words like "culpability," "excuse," "guilt," and the like. (76, 84, 97). For example: "The basic function of the concept [of voluntariness] is to inform the appraisal of individual conduct and in particular the assessment of innocence or guilt" (76). While we should not read "voluntary" and "morally responsible" as exactly co-extensive, they don't depart meaningfully for our purposes here.

to justify a high-stakes sacrifice, Rodong would have served under duress, and therefore involuntarily, and therefore without responsibility. Suppose, for example, his uncle threatened to have him and his parents shot. Or, if you still think resistance here is morally required, you can easily add to the tally of dead until you think the threat constitutes a sufficient basis for a duress plea. Here, the threat simply isn't sufficiently strong. Though the threat of losing one's university dream isn't insignificant, the consequences of joining such an evil organization don't allow for that threat to undercut the voluntariness of his action. Quite plausibly, Rodong is morally obliged to simply accept the loss.

2.4.4:The Nature of the Offense:

Here is Hyamn's line of reasoning in rape cases, reapplied to this case:

1) An act, like enlisting in the secret police, is voluntary if it is not the result of ignorance or duress/compulsion.

2) The ignorance excuse does not apply to Rodong; by stipulation.

3) If the stakes are high enough, and the threat relatively low, then the duress excuse does not apply.

4) Therefore Rodong's enlistment was voluntary (1-3).

5) Since Rodong enlisted voluntarily, he was not unjustly conscripted by Hyunbin (4).

6) If Hyunbin did not unjustly conscript Rodong here, he is guilty of nothing.

7) Rodong did nothing wrong (4, 5).

This argument has the same form as the reductio Hyman uses against volitionists. In his argument, imaginary volitionists are forced to say that women who accede to sex under threat are not rape victims, since they accede and (by stipulation) could simply not have acceded and born the whatever they were threatened with. But, says Hyman, this implies an absurd result, so we can jettison volitionism.

The scenario above shows that Hyman's own view has roughly the same problem. In that scenario, Rodong behaves badly. However Rodong's bad behavior does not imply that Hyunbin did not also behave badly. In this case, Hyunbin committed some kind of evil act concerning conscription, but the point generalizes to all crimes of force. Though victims may meet the minimum threshold for morally blameworthy action, this fact alone does not imply the blamelessness of their victimizers.

There are at least two ways of characterizing Hyunbin's offense. The stronger characterization is that he committed an act of unjust conscription. The upside of this view is that it preserves the seriousness of what Hyunbin did. Analogously, someone who threatens e.g. reputational damage to a woman (who for this reason agrees to sex), counts as a rapist. The downside of the view is that it does not preserve a tidy definition of conscription, where "conscription" means the involuntary enrolment of a person into a violent force. This is to reject premise (5) above.

If we wish to preserve that tidy definition, we might say that while Hyunbin did not conscript his nephew, he did engage in what we might call "military extortion," where military extortion is the act of immorally threatening someone into joining a violent group. This is to reject premise (6) above.

We could call the first option a "perpetrator-first" labelling strategy and the second a "victim-first" labelling strategy. I think the first option is preferable because it does less damage to real-world-based judgments about conscriptors, rapists, and the like. Hyunbin would be seriously punished in a just world for his actions, and the thing he would be punished for is usually called "conscription." "Voluntariness," unlike "conscription" or "extortion" or "rape" is a thoroughly philosophical concept, and it is *ceteris paribus* better to confine surprising definitional contortions to philosophical concepts.

That is not a knock down argument, but it does not need to be. Relatively little hangs on how we resolve this largely verbal dilemma, at least for present purposes. What matters is that Hyman's central argument, a *reductio* against volitionists, has an easy *tu quoque* response: his own view is committed to exactly the same kind of awkward categorizations.

The same point can be made a bit more strongly: Hyman's defenders should acknowledge that at least one of (5) and (6) above is false, since (7) is an absurd conclusion and (1-4) just are

Hyman's theory. But if either of those are false here, they are also false when Hyman applies them to volitionist theories of moral responsibility.

Hyman's view, like Hieronymi's, is incisive and clever. Also like Hieronymi's, it starts to break down under rigorous examination. Both fail because they skate over conceptual work about intention, choice, and voluntariness. Volitionist orthodoxy, insofar as it still is orthodoxy, was never vulnerable to end-runs of this sort—not, at least, if we look closely enough.

2.5: Conclusion:

This concludes the first part of the thesis. In this chapter, I have defended a volitionist account of moral responsibility from two worthy critics of the view. More could be said about how these deliberations might help refine volitionism. In particular, an account of the relationship between choice and intention would be welcome, for they would allow us to get to the heart of what volitionists want and help situate worries about externalities, negligence, and discretion. Though I do not attempt to do so, I believe the above motivates such a project. For present purposes, however, the volitionism survives.

It is thus of paramount importance to choose to learn what must be learned. Failing to do so is morally hazardous, for, if I am right, akrasia about our investigations and epistemic virtues lies at the very root of much culpability. There is tremendous moral value in getting to moral truth, for that, coupled with the grit to live the moral life, is the best we can do. This much is the claim ethics makes on epistemology.

In what follows, I inquire into epistemology's own axiology.[47] Above, I attempt to answer unanswered questions on the epistemic components of moral responsibility. Below, I cover ground even less trodden. There are, to my knowledge, almost no recent attempts to grapple with what is epistemically valuable to know. As I said in the introduction, there is much debate about which epistemic *states* are fundamentally valuable. Almost nothing is said about epistemic *content*. I hope to lay some groundwork for discussion of epistemic content in the following way: First I examine, in a

[47] The phrasing is a reference to Addler's (2002) superb book, *Belief's Own Ethics*, on doxastic norms.

necessarily cursory way, the work of Plato, Aquinas, and Hume on the subject. Second, I review what has been said about the intersection of epistemic state and content, and conclude that, sadly for Platonists, there is very little. Third, I lay the groundwork for a Thomistic-style approach, according to which epistemic value is both morally freighted in its own right and dependent on a certain metaphysical picture of the world. Finally, I attempt a charitable strengthening of the Humean view, which holds that curiosity specifically and practicality generally constitute epistemic value.

Chapter 3: Pre-Modern Conceptions of Epistemic Value:

Come now, I shall tell—and convey home the tale once you have heard—
just which ways of inquiry alone there are for understanding:
the one, that [it] is and that [it] is not not to be,
is the path of conviction, for it attends upon true reality,
but the other, that [it] is not and that [it] must not be,
this, I tell you, is a path wholly without report:
for neither could you apprehend what is not, for it is not to be accomplished,
nor could you indicate it. (Parmenides, Fragment 2)

3.1: Introduction:

In this section I seek to give the first brief history of pre-modern and early modern thought on valuable epistemic content. Such a scope is far too large for rigorous exegetical work, and such work is better left to proper historians of philosophy anyway. What I do hope to establish is a broader framework for thinking about epistemic value than is present in contemporary discourse. Whether the views I ascribe to Plato, Aquinas, or Hume are really theirs is not a question I hope my reader dwells on much. On one matter I am confident: In the pre-modern world there was a greater tendency to unite the methodological, axiological, and ontic with the epistemic. To ask the right sort of question, to find the logical structure, to secure truth, and to access the depths of reality were in some ways the same endeavor, particularly the further back we go.

If this seems irritatingly obscure, consider the present debate about epistemic value. With a few exceptions, most philosophers are chiefly concerned with finding the sort of epistemic *state* to ground epistemic value. Rivers of ink, to which I will add my trickle, have been spilt debating whether true belief, knowledge, understanding, warrant, practical utility, or some combination thereof ground epistemic value. Philosophers are very keen to know which ones, if any, ground what is valuable in the others. For our philosophical forebears, questions about what sort of questions to ask, about what sort of answer counts, about the content of what is known, and others are integral to

understanding matters of epistemic state; all bear some relation to all. Whether true belief caries the same value as knowledge is one question. Other important ones are whether knowledge of distal causal origins is of more value than knowledge of proximate ones, or whether knowledge of causes is of more value than knowledge of particulars, or whether true belief about God is of more value than true belief about insects.

That some ancients and medievals had unified theories of epistemic value is, I hope to show, relatively clear. The more difficult question concerns which aspects of these unified theories are fundamental. For example, if it turns out that understanding is more valuable than true belief, and it also turns out that understanding is understanding of causes, then we are left wondering whether it is the state (understanding) or the content (propositions about causes) or the referent (the fundamental causes of the world), (or some combination thereof!) that generates this value. It is a difficult question that will be addressed more in the next chapter. As such, I turn to the distant past not because I have any commitment to, say, the theory of the Forms, but because the distant past seemingly provides all the help there is to be had on these questions.

3.2: Plato:

Modern philosophers tend to think of epistemology and metaphysics as more-or-less separate disciplines, but the two were once much more entangled. Historians of philosophy long held that Plato, both the first and the best known to blur the two, thought the most important portions of reality are those that could be known in the strongest sense of the term, and that what was knowable was whatever was most real.[48] As White somewhat cautiously puts it, "[Plato's] views about what there is are largely controlled by ideas about how knowledge can be accounted for, and his thinking about what knowledge is takes its character from convictions about what there is that is knowable" (1992, 277).

[48] In this section, I will focus on Plato's middle period. Whether this period represents a development from his earlier views is unknown to me.

But there is no longer a firm consensus about this sort of metaphysical/epistemic blurring in Plato. I will first sketch the traditional "two worlds" interpretation of Plato, according to which epistemic hierarchies correspond to metaphysical hierarchies, such that that which can be known in the strongest sense is that which is most real. I will then attempt to defend it from Gail Fine's (1978, 1990) influential reductionist critique. On her view, or at least on a version of it, there is only one level of reality, and it has no particular epistemic state pertaining exclusively to it.

The following focusses on *Republic* V. Though there is more to say about other sections of the *Republic,* the *Phaedo*, and other dialogues, since *Republic* V is a primary locus of contemporary debate on the pertinent areas of Plato scholarship, I will follow suit. Now, in *Republic* V, Plato argues that knowledge is of what is and belief is over what is and is not. The trouble is that it isn't obvious what the meaning of ἐστί (translated, "is") is. Fine provides a menu, which I have somewhat simplified below:

1) Knowledge (or belief, or whatever other state) is set over what is in the *existential* sense of the word. Knowledge, here, is set over whatever exists. We know something when we know some (non-semantic) object that (really, fully) exists.[49] Correspondingly, belief is set over what partially exists, or is less real. We believe something when we believe something about a (non-semantic) object that partially exists, or exists somehow deficiently.

2) Knowledge is set over what is in the *predicative* sense. Here, it is set over what is F, where F is some suitable philosophical predicate (e.g. beautiful, small, human, etc.). We know something when we know some (non-semantic) object that is fully F. Correspondingly, belief is set over what is partly F. We believe something when we believe something about a (non-semantic) object that is less-than-fully F.

[49] A reader points out an ambiguity here. When I speak of knowing some object, I do not mean anything Husserlian about what is in my mind. Rather, I mean to speak of knowledge-about. The object of my knowledge, in this sense of "object" is something in the world.

3) Knowledge is set over what is in the *veridical* sense. This reading differs from the other two in two ways. First, here we should read the "something" in "whoever knows something" as concerning the content of the intellect, not an object out in the world. In contemporary parlance, we might call that content a proposition. Second, we should read "is" as set over whatever is true. Belief, correspondingly, will be set over what is and is not true.

    a. This might be construed to read that knowledge is over the set of true propositions,[50] while belief is set over both the true or false ones.

    b. The other construal is that knowledge is set over what is completely true, while belief is set over what is partially true.

We can take (1) and (2) to represent two versions of the traditional "two worlds" ("TW") view. When I know, I know something in the world of the forms. When I believe, I believe something in the world of the senses. (1) holds that there are degrees of existence ("DE"[51]); the forms exist fully, the sensibles partially. (2) holds that there are degrees of reality or being ("DR"); both forms and sensibles exist, but on different levels of reality or being. The difference between DE and DR is subtle, but both views ascribe to Plato the view that the world is divided, and that our epistemic capacities are correspondingly divided. (3) is reductivist. On (3a), which is Fine's view ("T"), Plato thinks we can perfectly well have beliefs about the forms, as well as knowledge of the sensible particulars. Rather, the important distinction is merely that knowledge is of what is *true*, while belief can be true or false. (3b) represents a degrees of truth reading ("DT"), a reading which will be easily accessible to readers familiar with contemporary debates about vagueness, non-Classical logics, and the like. On it, known propositions are true *simpliciter*, while believed propositions are neither fully true nor fully false.

[50] I do not mean to imply that Plato believes in propositions as such. I have no position on that question. As such, the reader is free to substitute "statements," or even "things."
[51] These labels have been taken from Baltzly (1997: 239-40). See for a helpful and more thorough account of Fine's taxonomy, to which I am here indebted.

Fine's taxonomy is not simply useful for Plato scholarship. It will provide a good tool for talking about how epistemic value relates to epistemic content in general. It will also come in handy in the discussion of Aquinas' views further on (though his views are not obviously identical with any of the options listed above). For now, however, I return to Plato.

A very short summary of how the argument in *Republic* V has been traditionally understood—Fine's target—goes something like this: Plato's aim is to convince those who love sights and sounds that they will never succeed in gaining knowledge from the sensible world. Knowledge, rather, is only of the eternal, unchangeably perfect Forms.[52] On the traditional account, the sensible world is stunted and incapable of generating knowledge. While it is suffused with being, it is only partly so, so its being is insufficient for knowledge of any kind. Sight-lovers errantly hold that they can look at the world around them and pick out those features available to sense that make objects beautiful, for instance. For example, they might think that the bright colors make something beautiful. But, says Plato, such coloring might be beautiful or ugly, depending on where it adheres. Bright blue, beautiful on a parrot or a sari, might be very ugly on an Oxford college. Beauty *as such* cannot be in any of them completely, for none of them are completely beautiful. Perhaps none of them even completely exist.

Friends of the sensible world, then, seek to understand what is and is not. On the traditional view, this "what is and is not" should be understood to refer to objects rather than to propositional content. As we have said, the tradition is split on how to understand this claim. DE claims that sensibles are not fully extant, while DR claims that sensibles are, for example, not fully beautiful (indeed, they aren't fully anything else either, and as such constitute a lower level of reality).[53]

Fine's classic paper (1978) argues that this traditional reading is mistaken. On her reading, Plato aims to start the argument with uncontroversial premises the sight-lovers would (and should)

---

[52] Importantly, even Fine admits some part of this. On her view, while knowledge need not end with the forms, it at least begins there. Second, on some readings old and new, we ought not think Plato completely confines knowledge to the forms. It might be that we can have knowledge, or something like it, of things similar to the forms.

[53] DR's defenders number Annas (1981, 209-25), Vlastos (1965), and many others. DE seems to become very unpopular in the last half-century.

accept. She means for Plato's crucial claim about knowledge and belief to be read according to T. So, roughly, that knowledge is of true propositions, while belief is sometimes of true ones (and sometimes of false ones). This reading holds that the questions are not about the objects that belief and knowledge concern, but their propositional content. Her defense of this move is three-fold. First, she thinks that these are genuinely uncontroversial premises—after all, Plato does explicitly frame them as such when he says that he will argue in a way acceptable to the sight-lovers, without beginning anywhere unacceptable to them. Relatedly, she argues that it would be strange to expect the sight-lovers, who believe they have knowledge, to so readily accept what looks like a Trojan Horse designed to smuggle in the forms. They have no reason to be so generous. Second, TW is (to her mind, anyway) strange and implausible, and we ought to be suspicious of it from the outset on those grounds. Third, this construal makes Plato's next bit of argumentation valid, and we ought to favor valid interpretations of great philosophy for obvious reasons.

If, then, we accept with Fine that Plato has only asked the sight-lover to accept (1) that knowledge is veridical, and (2) that the belief may or may not be, then Plato's subsequent argument in 477c is unobjectionable. There he argues that if faculties are distinct, they are distinct *both* in virtue of what they do and in virtue of that which they act upon. Now, on the traditional view, the faculty of knowledge and that of belief act upon different objects-in-the-world—for example the real and the less-than-fully-real—rather than merely on different sets of propositions. Fine thinks this analysis of powers is implausible. She offers the examples of husbandry and butchery. These faculties are obviously distinct, but seem to be set over the same object (namely, animals).[54] Rather, for Fine, only some powers have the structure Plato appeals to. If, following Fine, we read Plato as only claiming (T) initially, then we have such a case: The operation of knowledge takes as its object the set of true propositions; which is a subset of the set of propositions generally, while opinion's object is the different (though overlapping) set of all propositions. Fine takes it that Plato is referring to the sort of powers her interpretation of the initial premises picks out. The power of knowledge is something like

[54] Note that not all practices are distinguished in this way. Thievery, for instance, is distinguished from moving things around by ownership. It is not, then, obvious that Fine has a right to use this model for knowledge.

infallibility, insofar as it ranges over only the true. The power of belief is not infallible, and as such has a different object. The idea is to confine Plato's general-sounding (and implausible) claims about powers to those powers within the context of the dialogue; if we do as much, we don't have to worry about the total overlap husbandry and butchery have.

We now arrive at the central part of the argument regarding beauty.[55] Recall that the traditional view briefly alluded to above ascribes to the initial premises about the scope of knowledge and belief a predicative or existential reading of "ἐστί." On it, this step of the argument is predicative. Here, Plato argues that each of "the many" both is and is not. On the predicative sense of "is" at hand, we are to understand that each of the manifold things in the sensible world is both F and not-F (where F is any predicate—"beautiful," for example). Surprisingly, Fine agrees with the predicative sense here, though her argument is very intricate, for she must justify the claim that Plato shifts senses of ἐστί mid-argument, before switching back again, all without signposting it to his reader. The interpretation she ends up endorsing is that Socrates's reference to "the many" does not pick out the objects in the sensible world, but rather sensible properties. Bright colors, for instance, both are and are not beautiful, depending on where they are instantiated. Definitions, ever important for Plato, are thus beyond the reach of the sight-lovers. For whenever they define e.g. the beautiful with reference to the sensible they fail to get at the one essence shared by all members of the set they are trying to define. Not all beautiful things will be brightly colored, so it cannot figure into the definition. So important are definitions that without them, knowledge is beyond reach.

Finally, Fine argues that the final significant premise ought to be understood according the veridical sense of ἐστί. Namely, some of the sight-lovers' beliefs (νόμιμα) about the many F's are true, and some are false. As such, they have belief, but mere belief. If this reading is correct, we need not posit TW, since the claim that knowledge is only ever of what is boils down to the claim that knowledge is factive.

[55] For a clear summary and analysis of Fine's thought on this point, to which I am here indebted, see Baltzly (242).

Fine is presenting her opponents with a kind of dilemma. One horn is to read the first premise (about knowledge being only over what is) veridically, in which case any subsequent argument that reads knowledge in the predicative or existential sense is invalid. But Plato is a great philosopher, and those are not prone to giving such obviously invalid arguments. The second horn is to read the first premise predicatively or existentially, but then we have no reason to suppose the sight-lovers would be inclined to accept it. Since Socrates explicitly tells us he wants a premise anyone would accept (specifically them!), then this is not likely either. As such, we should read the whole thing veridically, and thereby jettison any wild notions about TW present in DE and DR.

The single predicative premise does not amount to equivocation, for it concerns only the basis of the sight-lover's beliefs, not his beliefs themselves. The source of his beliefs is indeed what (predicatively) is and is not. The beliefs themselves are and are not in the veridical sense. Second, it does not entail the Two Worlds theory, for what might be believed by one person can be known by another. And since the epistemic worlds are united, so too must the metaphysical ones. In this way Fine deflates the doctrine of degrees of being (or existence, or reality). We need not try to wrap our heads around the mystifying notion that the sensible world both is and is not, for all Plato means to say is that the contents of belief can range from the true to the false.

Fine's paper is an elegant and compelling piece of philosophy. Nevertheless, one worry is that Fine's view is in one way blinkered by a modern analytic mindset. Her method is simple: She takes the known senses of "ἐστί" and picks the one that would be most palatable to Plato's imagined contemporary analytic interlocutor. But this is the kind of move that MacIntyre, for one, would treat with suspicion. Fine's perspective, like all of ours, is that of an outsider looking in. It seems narrow-mindedly Anglo-centrist to read Plato by placing each use of "ἐστί" into Fine's schema above. Certainly as an exercise it is very helpful in clarifying the Greek, but it seems quite rigid to insist that Plato never simply uses the word with no particular sense in mind, or with more than one sense. It would not be particularly strange, for example, if his intended sense evolves as the dialogue unfolds. This sort of looseness might be foreign to the modern analytic philosopher, and her own careful attention to clarity might lead her to prematurely dismiss this possibility.

This line of reasoning should not be over-extended. It remains true that, whatever Plato meant, if he uses different senses of ἐστί in different premises, the argument is invalid. This does indeed count as mark against such a reading of Plato. However, it is also worth noting that some of Fine's opponents, Vlastos (1965) and Annas (1981, 209-15), are aware of this concern, and provide other compelling readings.

Also of MacIntyrian concern is Fine's rather bold view that the claims that we do know something of the sensible world or that knowledge is factive,[56] certainly held by almost every present philosopher, would not seem at all strange to people as temporally distant as fifth century Athenians. Indeed, plenty of people today seem to have differing intuitions about the nature of knowledge. Distinctions across cultures are particularly relevant: Experimental philosophers have found East Asians, for example, to have (in general) radically different responses to Gettier cases than Westerners.[57] That we find her version of Plato's argument more compelling is of limited utility, then. In the same way, it should not be surprising that Fine finds the Two Worlds ("TW") theory outlandish—even absurd—but that it nonetheless be Plato's view.

We cannot assume blithely that Plato's Socrates is above manipulating his opponents with bad arguments. Some scholars regard Straussian readings of the *Republic*, of which Allan Bloom's is the best known,[58] as slightly eccentric and outside the respectable mainstream. However, many of the core doctrines of Straussianism *viz.* the untrustworthiness of Plato's Socrates, his irony, and the subordination of his arguments to his philosophical project are within a range of acceptable views. As such, it is by no means clear that we ought to take Plato to be operating according to the principle of charity when Socrates says that we should "pacify [the sight-lover] and gently persuade him while concealing that the is not in his right mind." As Baltzly points out, Plato does not promise a fair fight. His opponent, after all, "is not his right mind." Indeed, he is not even willing, and possibly not even able, to follow a thorough argument (475d). We should not therefore be surprised if Plato exploits

[56] As an aside, suppose it *was* obvious that knowledge is factive. If so, it's at least somewhat curious that Plato spends significant time on the matter.
[57] Weinberg, Nichols, and Stich (2001). Whether claims like this survive replication is dubious. Since so little is yet known, I admit this argument is somewhat tendentious.
[58] Bloom (1968)

ambiguity in "ἐστί" in order to gently push his reader towards the doctrine of the Forms. In fact, Plato may simply be giving his own disciples a road map of sorts in the style of Maimonides,[59] drawn to help to them be subtle pedagogues like himself. The sight-lovers discussion precedes the full development of the doctrine of the forms perhaps because Plato wants to show how to gently entice those opposed to his most radical doctrines. As I hope to show, because the sight-lover is incapable of even starting philosophy, and because he is committed to a highly dangerous (but manipulable!) philosophy-manqué, then he must be tricked into accepting some basic precepts of Plato's method of inquiry.

Something like this seems to be Dirk Baltzly's view. Recall the initial setting of the passage. On a traditional reading, Plato begins by asking the sight-lovers to accept that knowledge is over what is, where "what is" is what is completely F (or exists completely). The sight-lovers might not be persuaded by such a claim, for they think their kind of loose and meandering discourses get at e.g. beauty perfectly well, without appeal to universal definitions. As such, the sight-lovers lack the foundational commitments necessary to Plato's approach to metaphysical questions. Concerning *Republic* V, "[Plato's] argument seeks to show that if one accepts certain constraints on answers to 'What is it?' questions, one must also accept an ontology of the forms" (243). But this will not be possible for the sight-lovers to accept without some subterfuge. Given how different they are in their fundamental commitments from the true lovers of wisdom, some degree of begging the question is necessary. As such, we should have no problem reading Plato's initial premises as Trojan horses smuggling in the forms through a predicative sense of ἐστί.

Balztly understands the sight-lovers to be dilettantes. Those whose flighty excitement is stimulated by a new gallery opening or a Woody Allen premier (his examples, 247) do not really want to come to an understanding of beauty or goodness so much as they want to discuss something less universal; the way that color interacts with shape in that painting, for instance, or whether such-and-such a character ought to have broken confidence. Passion they love; counterexamples they find

[59] Friedlander (1904, 11).

stifling and beneath the mystic dignity of the thing they have experienced. Philosophers, conversely, like counterexamples, universal principles, ruining films, and being clinically boring at gallery openings. For this reason, they are less prone to mistake a tokening of goodness as sufficient grounds for an understanding of goodness, nor a likeness of beauty as beauty *simpliciter*.

For Plato, knowing beauty requires access to what Baltzly calls its "logical cause," namely, that thing by virtue of which it is beautiful and by virtue of which all beautiful things are beautiful. This assumption is crucial to understanding how Plato's foundational methodological commitments (asking for definitions and throwing up counterexamples) is related, in some as yet unspecified way, to his ontology and epistemology. In modern terms, the sort of epistemic state Plato seems to require might be closer to understanding than to knowledge. Very often we speak of knowing something that we would not claim to understand; for Plato, this is (roughly-speaking) not the sort of epistemic labor worth the sweat and tears of dialectical reasoning. The "by virtue of" relation within logical cause is notoriously difficult to sort out, but what is clear is that the logical cause (by virtue of which X is F) is in some very important way *one and single* (249). Of course, while the logical cause of something's being beautiful is one and single, there are many beautiful things. But this is not to say that there are many reasons why each is beautiful, for if there were, none would be-completely beautiful (249-50).

This singularity requirement is thus intertwined with the second Platonic presupposition: the self-predication requirement, qua complete-being. "If x is that by virtue of which all F things are F then it is not the case that any sentence of the form 'x is not F πρός ψ' is true; e.g. if gold is the logical cause of beauty, then it cannot be true that gold is not beautiful or fine in relation to the function of a soup spoon" (250). Rather, it seems that all gold things would be beautiful. Of course, many statements about beauty will be true without reference to the logical cause. The sight-lover may say that texturing of St. Basil's Cathedral is beautiful, and he would be right. But if that texturing would not be beautiful πρός the Pantheon, or Helen of Troy's face, then it does not belong to the Socratic definition of beauty. That beautiful texturing (on St. Basil's) is beautiful, but the texturing itself is not, at least in the sense of complete-being.

The sight-lovers should not be understood to be searching for the logical cause of καλόν according to a different set of presuppositions. Rather, says Baltzly, "their lack of enthusiasm for philosophy as Plato understands it makes it *as if* they rejected these pre-theoretical starting points for Socratic questions" (251). Rawls' (1971, 432) famous example of the grass-counter concerns someone unable to take mathematics seriously, despite ability on his part, because of an unquenchable desire to know how many blades of grass are in his yard. We would suspect that the way to convince this person to return to mathematics would not involve taking grass-counting seriously. Likewise, the sight-lover's pathology precludes him from the arena. If Socrates can trick him by equivocation into delving into philosophy, all the better; especially, as the Straussians might emphasize, if there is some risk that he could get into politics. That could be even more disastrous than allowing the grass-counter to assign accountancy students' curricula. For, while the grass-counter is an obsessive, the sight-lover is a dilettante, and they are far more unsystematic, unpredictable, and chaotic, while being equally stubborn.

None of this is to say that there is nothing in Fine's project that is worth taking seriously. Rather, it seems that attention to Plato's methodology, his focus on metaphysical definitions, and his concern with the causal relations between sensible objects and the forms in which they participate all point to the traditional, orthodox TW view. At any rate, I will take it that such a view is at least defensible enough to be assumed. What, then, can we make of it?

Recall what Plato's methodology is. First, in order to make a judgment (an ethical one, say), we first need to know whether the action (or way of life, or city-state, etc.) counts as F, where F is some relevant category like the pious, the beautiful, or the just. Next, we need to find out what e.g. the pious is through dialectical reasoning and counterexample-testing. Once we have a working definition, we have either a form or something fairly close to it, for these definitions will have to pass the singularity and self-predication-by-complete-being tests. There will be one thing in virtue of which pious things are pious, that thing will be completely pious, and insofar as it is present anywhere, it will always be pious. Naturally, this is not to say that everything it participates in will be

pious. An action might be pious qua filial duty, but impious qua divine obligation. But insofar as any action participates in the Pious, it is pious.

This sort of methodology might be called geometrical. Baltzly suspects that "Plato, and perhaps Socrates, thought that these conditions were just obviously reasonable for the possibility of properly scientific answers: there are not many explanations/causes for why a figure is a triangle nor is there every any respect in which a three-sided plane figure is not a triangle" (252). The lovers of sights and sounds, who travel from festival to festival feeding their curiosity about beautiful things, would be unlikely to accept this sort of method, so it is unsurprising that Plato is less than explicit about it.

Obviously, Plato's epistemology is both too nuanced and too large to be easily encapsulated. But a low-resolution summary of it reveals a kind of epistemic matrix:

1) Epistemic Goal-State: The ideal epistemic state to be obtained is knowledge, which is a durable good, worthy of the very considerable effort.

2) Ontic Goal-Content: The ideal epistemic content is of the forms, which are the most real.

3) Methodology: The ideal way to go about searching for worthy epistemic goods is by Socratic questioning; that is, by posing counterexamples to hone in on a definition.

4) Success conditions: We have attained something epistemically worthy when we know the that which is completely F, and that which (in some way) causes F-ness in all F things.

The language of epistemic value is contemporary enough that it doesn't translate easily into 5th century Athenian thought, though I also suspect it is loose enough that it isn't totally senseless to talk about Platonic conceptions of epistemic value. If this is right, we are allowed to query the relationship between these four. Which one (or several) are fundamental goals for the epistemic enterprise?

I think we can rule out any interpretation in which (3) is fundamental. One could perhaps make a case that human beings are prone to ask certain kinds of questions, or to engage in certain kinds of peculiar methodological quests, and that any value at arriving at an answer is purely a

function of, say, sustaining that methodological enterprise. But this is bizarre and unmotivated. Surely (3) is tailored to the acquisition of something, not the other way around. Something similar must go for (4). Nonetheless, methodology is worth taking seriously. First, it helps show what kinds of knowledge are the ends of proper investigation. This reveals the primacy of metaphysical (or perhaps ontological) knowledge. For Plato, in order to know, we must understand what something *is*. True propositions, *pace* Fine, are not the highest epistemic goods in themselves; rather, only the subset about the Forms (and their cousins, the ideas) are what can be known, for only they are fully real.[60] Even when we correctly believe something, we do so with reference to them. When I believe Helen is beautiful, I believe truly insofar as Helen bears the right sort of relationship with the beautiful. The forms are the beautiful-makers for Helen, and perhaps something at least akin to the truthmakers for propositions about her beauty.

As far as I can see, (3) and (4) do not reveal which of (1), (2), or both is the primary locus of epistemic value. But they can reveal something about how the quest for wisdom seems to be directed towards foundational reality. It seems that the quest for understanding we gravitate towards involves asking lots of "why" questions. This, of course, is not just true for us, but for Plato and, for that matter, the sight-lovers. But the really important kinds of knowledge seem to be those that get us towards the right definitions; towards what is completely F, such that no sentence of the form "such-and-such is not F in some respect" is ever true. Contemporary analytic philosophy tends to think of the acquisition of wisdom in terms of chains of why questions. Because we want to understand, we want to know why things are the way they are, and the more fundamental the answer, the deeper the truths we get at. The necessary truths or brute facts that lie at the ends of those chains are, importantly, still propositional. For Plato, the ultimate truths are forms, and we know them by

[60] In what sense can we say that we might know the forms, rather than know about the forms? Plato does not talk about propositions. He talks about knowing what is, and (according to my interpretation) that may mean having some sort of acquaintance with the form at hand, rather than mere knowledge *about* it. Further on, we will see how Aquinas thinks knowing God is far more important than knowing the lesser (though still important) truths about God. Given the geometrical, definition-focused nature of the Platonic method, I do not think Plato is so concerned with the distinction.

producing and understanding correct definitions. “What” questions are, in the end, most fundamental, for the ultimate objects of knowledge are themselves entities.

There is no question that Plato’s doctrine is outlandish in modern eyes. But it is not so strange a claim that the most valuable sort of epistemic content is the sort obtained by asking deep questions. It is also not so strange that the kind of secure understanding obtained by pursuing those deep questions is relatively secure, and that it is in some epistemically significant way superior to mere true belief about those questions. Further, it might be that we might get at deeper questions by asking “what is F” in the right way. That is, by trying to get at the cause of the phenomenon we are interrogating, and then trying to get at the cause behind that, and the cause behind that (or perhaps, the cause beneath that), and so on. And, finally, as we Socratically progress through ever-deeper causes, it might be that we reach Being itself, or God, or the Fundamental truth. Something of all these suppositions is in Plato; something more is in the Scholastics.

3.3: The Theistic Shift:

The introduction of the Abrahamic religions to philosophy importantly alters the epistemological landscape in a way that can be summarized easily, albeit in a way that ignores some complexities:

1) If there is a (Theistic) God, knowing about God is important. In fact, God is the most important thing to know about.

2) However, God far exceeds our capacities for understanding. What knowledge we have of an omnipotent, omniscient, omnibenevolent, personal God is profoundly limited.

3) As such, that which is of highest epistemic value and that which we can know with the deepest understanding come apart.

The medieval philosopher I focus on, Thomas Aquinas, lays out the case in a canonical way. If he is right, then independent analysis of valuable epistemic content is necessary.

The transition was a gradual one. Platonism maintained its influence over the Western mind through the work of Augustine, but not without significant alterations. He spends much of the *Confessions* describing his tortured, faltering, and ultimately catastrophically unsuccessful attempts to achieve knowledge of the highest things through reason. Whereupon, famously, he makes an act of faith, and thereby discovers the wisdom that was eluding him. Augustine took himself to have gained something crucial to human flourishing and salvation, but, importantly, something that was as accessible to ordinary people as it was to a great mind like his own. Yet, upon accepting divine revelation, Augustine came to see that there was still a great deal of work for the philosopher to do. The mysteries of revelation, while known through faith, could now be explored through human reason profitably. Though full comprehension would be impossible even in Heaven, much of importance could be understood even now, albeit through a glass darkly. As his famous dictum goes: "Understanding (*intellectus*) is the reward of faith. Therefore seek not to understand that you may believe, but believe so that you may understand."[61] Some great part of Augustine's project thus became retooling philosophy (which, for him, more-or-less just was Plato and Plotinus) into a means of understanding Christian revelation. [62]

Yet Christianity is only compatible with Platonism to a point. Above, I argued that the traditional (two worlds) understanding of Platonism was the correct one: the highest things are those things which can be known with confidence and with understanding. Now, uncontroversially, God is the highest being in Christianity, and while some things about God can be known through philosophy, almost every Christian thinker[63] recognizes that much about God (say, his triune nature), cannot be fully comprehended philosophically. It may not be immediately apparent that this is at odds with Classical thought; after all, Plato knew that knowledge of the forms was extremely difficult to obtain. For most, both the One and *Noûs* were entirely elusive. However potential for conflict remains. First,

---

[61] This aphorism is helpfully compared to MacIntyrean thought on tradition. Some forms of understanding are unavailable to those outside the relevant tradition; they can only be understood, perhaps, "from within." Only once the hard work of immersing oneself in the tradition is complete can the justification (and reward) for doing so be perceived clearly. In such cases, one must resort to rhetoric to convince outsiders to do the requisite hard work of immersion.

[62] See Gilson (1938, 1-15) for a discussion of the anti-philosophical strand of medieval Christendom. See also (16-23) for a good introduction to Augustine's approach to philosophy.

[63] Anselm of Canterbury might count as an exception.

at least in Plato, the forms are *in principle* accessible to the sublunary human rational intellect. Second, they are *only* accessible to the loftiest operations of reason: to *episteme*. It is not entirely clear whether Augustine disagrees with the Christianized version of the first claim. *Some* understanding of the Trinity, for example, seems to be possible. The second claim, however, is totally antithetical to Christianity, for the ordinary faith of the uneducated and unphilosophical allows for a deep union of the intellect with God. And, whatever the faith of the common Christian may be, it certainly is not analogous to the rational demonstration of mathematics. Augustine, for all his energy, never managed to put this awkward conflict with Plato to bed.

3.3.1: Averoist Islamic Influence:

The medieval mind was drawn to no one solution exclusively. Many of Aquinas' views, to which we will turn soon, were shaped as a critical response to the fideistic "Tertullian" streak still present in Europe post-Augustine. But he was also in dialogue with Averroës (Ibn Rušd), a man whose commentaries on Aristotle he respected so much that he referred to him simply as "The Commentator." The Averroist solution to the problem of revelation is so Straussian that Strauss' Plato would blush. For Averroës, there were three kinds of people, with one kind of grasp of the truth appropriate to each. The common man was unable to grasp much beyond images; for him the Koran leaves tantalizing images of eternal corporeal pleasures should he live moderately, and terrifying punishments should he not.[64] The rules they live by do not amount to deep truth, but they give the masses some concretized version of it, which is the best anyone could have hoped for. For the philosopher, the mathematically precise demonstrations of classical philosophy are the work of a good Muslim, and knowledge is his reward. For the intermediate class of people, who can understand dialectical reasoning, though not precise, necessity-achieving demonstrations, the Koran gives theology. These people cannot attain real knowledge, but they are too intelligent for mere images to suffice. Some rational light is required, and probabilistic, dialectical arguments are within their grasp. So, revelation dresses up the basic doctrines given to the lowest class in the garb of the highest one,

[64] As Gilson somewhat jocundly puts it: "You will not civilize a tribe of Bedouins by teaching them metaphysics" (43). It bears stating that Averroist approaches are by no means the only Islamic view on the epistemology of religion—nor are they at all common in post-Medieval Islam.

and those who already believe the faith are able to talk to each other about why it is true. His view of this theological class bears some similarity to Augustine; but he has a far dimmer view of what theologians accomplish. Without theology, however, this class of people would be left without faith *and* without the real philosophy just beyond their ken; as such, while we ought not take theological claims too seriously, we ought to be very glad of its existence as a way of keeping the mid-witted out of trouble.[65]

Such a view, perhaps unsurprisingly, drew condemnations from Islamic theologians. When it made its way to Europe, many Christian theologians were similarly unimpressed. Many, however, took this view on board. How could it be that the intractable, mystical, even incomprehensible doctrines of the Incarnation or the Trinity were the subjects of knowledge superior to the demonstrations of Aristotle? Gilson distinguishes two distinct types of these Latin Averoists: Some were sincere Christians who upheld the truth of Christian revelation, but nonetheless maintained that some form of Aristotelianism was "the *necessary* results of philosophical speculation" (57). Alas, so much the worse for philosophy, they thought. Philosophy says what it says, and what it says must be true in some sense; yet, where it is at odds with revelation, we ought to believe revelation.[66] Others seemed to merely maintain Christianity publicly, with attitudes ranging from winks to snarls.[67]

This then, was the context Aquinas found himself in. Augustine had left unfinished the business of the epistemology of faith, and his project was beset by lingering fideism in the Church on one side, and the rising specter of Latin Averoism on the other. In the following section we will see how, for Aquinas, the highest kinds of knowledge and the highest parts of reality do not correspond in the same way as Plato, though there are many similarities. In both, the best things to know are those things according to which everything else may be known. This is to say, there is broad agreement about the sort of content that we ought to seek. The difference lies in the epistemic states involved.

---

[65] This summary is reliant on Gilson (38-53).
[66] Siger of Brabant, the most famous Latin Averoist, seems to be in this category. Dante, for one, was convinced of Siger's deep Christian faith (see *Paradiso* X).
[67] John of Jaudun, when publicly professing to believe in creation *ex nihilo*, added that "creation very seldom happens; there has never been but one, and that was a very long time ago." It does not take a very great leap of faith to see that faith may not be something he took very seriously.

For Aquinas, the sort of "value" that is most appropriately called "epistemic" consists in the perfection of the intellect by its union with truth. Since the most important *truth* to be known *is about* God (or, perhaps, the most important *thing* to know *is* God), the greatest perfection of the intellect lies in its union with God. However, we cannot know God through demonstration, but only through our faculty of common knowledge, whose justification at least begins in second-order justifications for belief. Our beatitude in this life, the perfection of our intellects, and what looks like an attractive candidate for the highest epistemic values is achieved not by a loft kind of epistemic state, but by a particularly precious sort of epistemic content: the truth about the highest things.

3.4: Aquinas:

Thomistic epistemology is a complex and nuanced affair spanning many texts. As before, I do not intend to provide more than a rough sketch of Aquinas' view of epistemic value, and as before, I will not attempt to integrate his whole corpus. My focus here will be on a few texts. First, on the *Summa Contra Gentiles*, which is one of Aquinas' more strictly philosophical[68] works, and one directed at the conversion of non-Catholics generally and Muslims specifically by making known "the truth that the Catholic faith professes (*Summa Contra Gentiles* I, ch 2, [2]):

> Now, the end of each thing is that which is intended by its first author or mover. But the first author and mover of the universe is an intellect, as will be later shown. The ultimate end of the universe must, therefore, be the good of an intellect. This good is truth. Truth must consequently be the ultimate end of the whole universe, and the consideration of the wise man aims principally at truth. So it is that, according to His own statement, divine Wisdom testifies that He has assumed flesh and come into the world in order to make the truth known: "For this was I born, and for this came I into the world, that I should give testimony to the truth" (John 18:37).
>
> The Philosopher himself establishes that first philosophy is the science of truth, not of any truth, but of that truth which is the origin of all truth, namely, which belongs to the first principle whereby all things are. The truth belonging to such a principle is, clearly, the source of all truth; for things have the same disposition in truth as in being (*SCG* I, ch. 1, [2]).

It may not come to as a surprise to the reader that God lies at this origin. He is the ultimate end of the whole universe, and contemplating God is the perfection of the human mind and the telos

[68] "The Muslims and the pagans accept neither the [Old Testament] nor the [New Testament]. We must, therefore, have recourse to the natural reason, to which all men are forced to give their assent" (*SCG* I, 2, [3]).

of human existence. Unfortunately, one kind of knowledge that comes from demonstrations, namely, Aristotle's understanding of "what the thing is" in the *Posterior Analytics*, is impossible with regard to God (at least in our present state). For, says Aquinas, "the human intellect is not able to reach a comprehension of the divine substance through its natural power" (*SCG* I, ch. 3, [3]). With respect to God's substance, this is something like a fundamental earthly limitation, and not one (simply) due to human sinfulness.[69] Aquinas is very explicit about the possibility of some kind of vision of the divine substance with the help of God later on (*SCG* III, ch. 51), but without miraculous intervention it is certainly impossible; with miraculous intervention it may still not be the case that anything like Platonic episteme is possible.

Still, in this life, some things the intellect can know in the strongest sense, through demonstration. For example, it can know that God exists through natural reason, unaided by revelation. Other things about God, like his triune nature, are beyond human understanding, though some sort of very modest understanding is possible. Why, then, ought a rational person to believe them? His answer is, in part, fairly straightforward.

3.4.1: Faith:

In the fourth chapter of the *Summa Contra Gentiles* Aquinas argues for two kinds of second-order reasoning towards the truth of revelation. We know from the fifth chapter that reason can never come into conflict with revelation, but this only shows that revelation is possibly true from the subjective human standpoint, not that it is true. One might think that what Averroes called dialectical argument would be the means of securing its truth, or at least its probable truth. But this is not what Aquinas argues. Dialectical reasoning, the classical route to justified strong opinion, does exist in Thomistic epistemology, but he does not appeal to it here. Nor is the even lower mode of conjecture (roughly, that one could easily take something to be true) obtain. Rather, one takes the truths of faith to be true through testimonial belief. He says more in the *De Veritate*:

[69] Adam and Eve, prior to their fall, did not see God fully ("in his essence"). That reward is for those who have reached a perfection beyond natural human perfection in heaven. See *DV* 18, 1.

> The essential object of faith is first truth. This should be understood from the following. Only that habit has the character of virtue whose act is always good. Otherwise, a virtue would not be the perfection of a power. Accordingly, since the act of our understanding is good because it considers the true, it must be impossible for a habit existing in the understanding to be a virtue unless it is such that by it one infallibly speaks the truth. For this reason opinion is not an intellectual virtue, whereas scientific knowledge and understanding of principles are, as is said in the *Ethics.*
>
> However, faith cannot thus stand as a virtue, deriving from the evidence of things, since it deals with things which do not appear. Consequently, it must derive this infallibility from its adherence to some testimony in which the truth is infallibly found. But, just as every created being of itself is empty and liable to fail, unless it is supported by uncreated being, so all created truth is liable to fail except in so far as it is regulated by uncreated truth. Hence, to assent to the testimony of a man or an angel would lead infallibly to the truth only in so far as we considered the testimony of God speaking in them. Consequently, faith, which is classified as a virtue, must surpass the truth of man's own understanding and thus make it embrace that truth which is in the divine knowledge (*DV*, 14, 8).

Common belief, for Aquinas, is believed in the absence of any kind of understanding. No first order evidence of the target belief is involved. It is simply taken on authority. Very importantly, it is taken on *God's authority*. This reliance on authority can be rational, if the agent in question has sufficient evidence of the reliability of the testimony in question,[70] but it is a very low epistemic state—perhaps lower than any other state than what he calls doubt, which is merely taking something to be possibly true without being believed. [71]

Unlike stronger sorts of states, which produce belief immediately upon cognition, common belief requires an act of the will. Common belief is a broad category, and need not concern faith that the revelations of the Christian God are true. When it does however, the believer is infused with grace that grants certainty and allows for deeper exploration of God and all reality. What it does not do, however, is raise faith to the level of demonstrative knowledge. It is held with certainty, but there is insufficient comprehension to be known the way mathematical proofs are known. It is "midway

---

[70] Aquinas gives a number of examples in the sixth chapter of *SCG* I. He appeals to miracles, to the success of early Christianity despite its ascetic attitude towards bodily pleasures and despite its unwillingness to conquer through violence. Miracles continue to be performed, but he does not see them as the primary evidence: "This wonderful conversion of the world to the Christian faith is the clearest witness of the signs given in the past; so that it is not necessary that they should be further repeated, since they appear most clearly in their effect. For it would be truly more wonderful than all signs if the world had been led by simple and humble men to believe such lofty truths, to accomplish such difficult actions, and to have such high hopes." In short, the argument is not just that one ought to trust the Church because of miracles, but because it has been far more successful than one would expect on a non-Christian picture.

[71] See De Haan (2015) for a very clear treatment of Aquinas' gradations of knowledge and their relation to the God.

between knowledge and opinion" (*Summa Theologiae* 2a2ae 1, 2). Like knowledge, it is a deep conviction, and one that is (at least in some cases) certain. Like opinion, it does not involve full understanding of its object.[72] At its most exalted, faith is "described as knowing God, [and as] an that by which the mind shares in God's own knowing."[73] At its lowest, faith is an epistemic state that involves so little understanding of what it believes that it is totally reliant on the testimony and authority of others.

In a moment I will say something in response to the obvious question regarding how and why the believer knows she has the right God. It is not strange to believe something on God's authority—but it is important to make sure God is actually saying it! In order to see the Thomistic answer, we need to understand first what faith in God is, and how it unites the mind to the heart and fountainhead of reality.

3.4.2: Views on Faith Falsely Attributed to Aquinas:

A) Suarez's

On one view, the God grants revelation to prophets and saints. These are written down in the form of scripture and doctrine, and carefully curated by the teaching authority of the Church. As the Cambridge UP commentary puts it, "believers accept these truths in the way understood by the Church, teaching under guaranteed guidance by the Spirit of truth" (193).[74] Analogously, we might think of the way we think of the scientific textbooks. We believe what they say because we trust the scientists who wrote them, and (more-or-less) trust that the

[72] I would note that none of this is to say that what is believed through faith in Aquinas does not count as knowledge as the word is understood today. The belief in question (for Aquinas) is believed, is true, and is justified. It may be that it is sufficiently justified, or sometimes sufficiently justified, to count as knowledge on a more contemporary understanding of the term. That said, it is not so apparent that it does not require some act of the will. This act of the will, what Aquinas calls an "inner assent," is some sort of act, but we can remain neutral on whether it is an intentional action to which doxastic voluntarists appeal. The matter is complex, and Aquinas' treatment of it would require more discussion than is appropriate for our purposes.

[73] See The CUP introduction to the commentary on the *Summa Theologica* (1911) on faith, xviii.

[74] See Latourelle (1968) for helpful work on Suarez's view. I note that this view is sometimes passed off as Aquinas', but that Suarez himself does not do as much.

views of the scientists have been preserved and transmitted effectively. On this view we reach God by first accepting faith, rather than the other way around.

B) Plantinga's[75]

On a second view, Aquinas is more-or-less an evidentialist. We have already seen Aquinas (*SCG*, I, 6) argue for the reasonableness of Christianity (in contrast to Islam) by pointing to evidence (the success of Christianity in the absence of military conquest, miracles, and so on). For Plantinga (1984, 46), Aquinas is arguing that "to believe in the mysteries of faith is not foolish… because we have *evidence for* the conclusion that God has proposed them for belief." He goes on to say that it would be foolish to believe without this evidence, and that Aquinas' considered view is that "It is just because we have evidence for these things that we are not irrational in accepting them."

Both views have some merit, and both have echoes in Thomistic thought. However, neither attend to a technical feature of Thomism: formal objecthood (*formalis ratio objecti*). Now, all human acts, capacities to act, powers, and habits of acting have both formal and material objects.[76] The material object of an act is that which an act acts upon. If I drink from my cup, the coffee is the material object of my drinking. This is not to say that the water is a material object independent of me, but it is insofar as I engage it.[77] The same relationship holds for other sorts of objects. If I see a window, the window is the material object of my seeing; if I believe in evolution, the truths about evolution are the material object of my belief; if I love my mother, my mother is the material object of my love. But saying that water, windows, evolution, and my mother are material objects does not tell us why I drink, see, believe, or love them. For this we need the formal object. Now, for Aquinas, the formal object, roughly speaking, is what draws me to material objects. Coffee is tasty, and therefore

[75] I am indebted to Siebert (2015) for collecting and criticizing various interpretations of Aquinas on belief here.
[76] Sometimes called formal "objective" or material "objective."
[77] See the CUP Commentary: "The very use of the term 'object' denominates things, realities, persons, in view of man's distinctive engagement with them. Apart from this perspective, they are not objects, they are just themselves. There are no material objects or formal objects 'out there'; there are things, there are persons" (182).

appears to me as good. The window is visibly extant, and therefore puts forward knowable truth—as does the theory of evolution. My mother is lovable, and this lovableness draws in my love:

> With respect to knowledge, it is the actuality of being itself that 'allows' knowledge and to which the knowledge is a reaction. In the case of appetition, it is the goodness, therefore the actual being, of the object that provokes the corresponding love, desire, hope. With respect to love in its most proper sense, love of a person, it is a response of mutuality, and active, affective union matching the self-giving of the other (CUP Commentary 182).

Like with the other theological virtues of hope and love, both the material and the formal object of faith is God, for it is by God that the mind can come to know God. Since God is beyond human comprehension, the kind of evidence that would allow the human mind to understand God is unavailable. What remains is God's grace that calls human beings to faith on God's own testimony. Anscombe, a modern Thomist, distinguishes between believing what a speaker says and believing the speaker. It may be that you happen to believe what I, a known liar, tell you, for perhaps you have other bits of evidence that I am telling the truth. But this is not the same as believing *me*. As Siebert points out, on the Plantingan construal, there would be no need to believe God when we can believe the evidence.[78] He compares faith in God to faith in a geometer. If my friend, a trustworthy geometer, tells me of the truth of some theorem, one which to him is obviously true and well within the widely-set limits of his comprehension, then I would believe him. In such a case, though our material objects are the same (in this case, the proposition), the formal object of his knowledge would be the evident truth of the theorem, whereas mine would be my friend. The knowledge of faith is most often directly compared to hearing, rather than seeing:

> Other things being equal sight is more certain than hearing; but if (the authority of) the person from whom we hear greatly surpasses that of the seer's sight, hearing is more certain than sight: thus a man of little science is more certain about what he hears on the authority of an expert in science, than about what is apparent to him according to his own reason: and much more is a man certain about what he hears from God, Who cannot be deceived, than about what he sees with his own reason, which can be mistaken (2a2ae 4, 8, ad2).

[78] Something similar goes for the Suarezian view. It is addressed in the following section.

It is then no anachronism to say that the most important kind of human knowledge, for Thomas, is second order authoritative (or testimonial), rather than inferential, properly basic, or (non-testimonial) evidential.

This may seem a rather pedantic subject, and to some degree it is. Its importance lies, however, in the way it underscores the extent to which faith in Aquinas is important not merely as an intellectual accomplishment, but as an interpersonal one—even a moral one. Faith is of course an intellectual thing (and Aquinas takes great pains to explain as much). But it is also a moral virtue second only to love. Faith allows the believer to unite herself with God as both formal and material object; that is, she both believes in God and she believes God, all at once. Indeed, it is the *only* way she can know God (at least in this life), given that the divine mysteries, while never demonstrably irrational, far exceed the limits of human understanding. When properly formed, Aquinas argues (2a2ae, 17, 8), the relationship with God is suffused with the love "the father and friend who speaks."[79] In this way it constitutes a deep trust, the value of which does not seem to be exhausted merely as intellectual accomplishment. This in turn qualifies the way we ought to understand the value of knowledge of God for Aquinas. While it is likely valuable as knowledge, as I will show below, it is not *merely* this. Whether it counts as primarily "epistemic" value is not obvious.

3.4.3: A Problem:

Before moving on, I should address the obvious question concerning the rationality of faith. The basic problem is the circularity of the Thomistic account. Suppose an unbeliever is told that the Catholic faith is true, and that he can know that by placing his trust in God. He might have questions he wants answered before he puts his trust in God. First and foremost, he will want to know if there is a God, and if there is, if that God is the same as the one being proposed. In the *Meditations*, Descartes is sensitive to this worry, noting that "although it is absolutely true that we must believe there is a God, because we are so taught in the Holy Scriptures, and, on the other hand, that we must believe the

[79] See CUP Commentary, 192.

Holy Scriptures because they come from God …, we nevertheless could not place this argument before infidels, who might accuse us of reasoning in a circle" (Haldane and Ross 1967: 133).

The obvious recourse is the Suarezian one: The believer believes that e.g. the teachings of the Catholic Church are true on the basis of evidence (testimonial or otherwise), and thereby comes to faith. After this, he can trust in God as needs be. However, this is not Aquinas' view, because it undercuts faith's nature as a theological virtue with its formal object as God. As I have already noted, if the act of faith is not an act with God as its object, then it is not the sort of faith he is interested in:

> Faith, in this conception, could remain a theological virtue—a virtue with God as its immediate term—only by either identifying the teaching Church with the first truth; or by maintaining that the acceptance of the contents of faith, its articles, is an immediate reaching of God himself (O'Brian 193).

That door, then, is closed to the Thomist. I am unsure how to respond to the objection, but part of the answer may have to do with the volitional aspect of faith. Faith is second-order testimonial knowledge, but it is enabled by further second order knowledge. Various kinds of evidence make it clear to the potential believer that faith *might well* be rational. On this much Plantinga is correct. There are numerous arguments throughout Aquinas' corpus that range from philosophical deductions about the existence of God to second-order arguments about the reliability of the testimony of the Church. It is then up to the individual's will to take the plunge into faith or to refrain (See the first chapter of *De Veritate*). Whether or not this sort of act may be meritorious, or blameworthy, or constitutive of some kind of doxastic voluntarism is a question for Thomists. I merely suggest it as a way out for the Thomist, and return to the epistemic questions that are our primary concern.

3.4.4: The Intellectual Value of Faith:

In Classical thought, as we have seen, the highest epistemic state was, broadly speaking, appropriate to the highest objects. This, Aquinas denies, albeit in a qualified way appealing to God's grace as the sole means by which man might be lifted up to some kind of knowledge of the divine nature. None of this directly shows that content is the primary source, nor even *a* source of epistemic

value (though it certainly seems to suggest as much). There are ways in which this could fail to be so. Here are two:

1) It may be that knowing God is valuable in the final sense, in that it brings about happiness, salvation, or other goods. But this does not imply that it is *epistemically* valuable.

2) It may be that knowing God is epistemically valuable in proportion to the lower epistemic state by which God is known. So, while knowing God is somewhat valuable, knowing other things through demonstration is more valuable.

In the style of Aquinas, I will take each point in turn. Now, the first objection is a difficult one for exegetical/historical reasons. It is not easy to specify what epistemic value is in contemporary analytic philosophy; it is even more difficult to say what it would mean for Aquinas to have a "theory of epistemic value." The notion of value itself is a nineteenth century innovation. But, as I mentioned at the outset, this chapter is not precise exegesis, but a mining expedition into the past. I will leave subtleties to the historians of philosophy. Nevertheless, we are not without some broadly appealing interpretations. The intellect is the epistemic faculty, and its good is likely to be related to that which is valuable for it, whatever "value" might come to.

In the third book of the *Summa Contra Gentiles*, Aquinas addresses this subject directly. There he argues that intellectual creatures achieve their ends in the understanding of God, and indeed of his substance. In a passage reminiscent of Plato, he argues that operations like understanding, which are the most perfect operations of the intellect, "are specified by their objects… [and] any one of these operations must be more perfect when its object is more perfect" (25, 3). As such, to understand the most perfect object is the most perfect kind of understanding. How are we to read "*intellectus*," here translated as "understanding"? While its precise nature is not obvious, we are told immediately that it does not consist in the kind of complete grasp characteristic of Plato's episteme or Aristotle's demonstrations:

> Of course, someone could say that the ultimate end of an intellectual substance consists, in fact, in understanding the best *intelligible* object—not that the best object of understanding

> for this or that particular intellectual substance is absolutely the best intelligible object, but that, the higher an intellectual substance is, the higher will *its* best object of understanding be. And so, perhaps the highest created intellectual substance may have what is absolutely best as its best intelligible object, and, consequently, its felicity[80] will consist in understanding God, but the felicity of any lower intellectual substance will lie in the understanding of some lower intelligible object, which is, however, the highest thing understood by it. Particularly would it seem true of the human intellect that its function is not to understand absolutely the best intelligible object, because of its weakness; indeed, it stands in relation to the knowing of the greatest intelligible object, "as the owl's eye is to the sunlight." (*SCG* III, ch. 25, [4], emphasis mine).

Here Aquinas posits that perhaps it is better to have a more complete understanding of something lower than the divine being, for such an understanding would be intelligible to it. This shows that the sort of understanding of God in which we find the fulfilment of our human teloi; namely, the understanding of God possible in heaven, where our intellects are infused with the divine light of grace, still does not consist in full understanding.

As such, complete understanding, of the sort definitive of demonstration, is not the fulfilment of the intellect. However, some blinkingly owlish form of understanding is. As such, we can answer the first objection. There we asked whether it was epistemic value (the good of the intellect) or some other kind of value at hand. It turns out that the end of human beings just is the good of human intellects, so the problem dissolves. We are also almost in a position to answer the second objection. There we asked whether the content of our knowledge of God was the primary generator of value, or whether it was merely the case that a diminuated form of knowledge was the locus of a diminutive kind of value. In response to this kind of worry, articulated in the passage above, Aquinas asserts that it is simply obvious "that the end of any intellectual substance, even the lowest, is to understand God" SCG III, ch 25, [4]). For, since everything in the universe tends towards God, something as noble as a human being ought to find fulfilment in understanding God, albeit through a glass darkly. "Besides," he adds, "a thing has the greatest desire for its ultimate end. Now, the human intellect has a greater desire, and love, and pleasure, in knowing divine matters than it has in the perfect knowledge of the

[80] By "felicity," sometimes translated "beatitude" we ought to understand the achievement of man's final end in heaven. The term can also be used relative to earthly life, insofar as things can be more felicitous insofar as they more closely resemble final perfection with God.

lowest things, even though it can grasp but little concerning divine things. So, the ultimate end of man is to understand God, in some fashion" (SCG III, ch 25, [7]).[81]

However, one thing must be conceded. One may read Aquinas as holding that intellectual ends are not ends in themselves, and are valuable only insofar as they contribute to happiness. Happiness, for Aquinas is man's final end. Knowing God is man's end *insofar* as it brings about man's happiness. So, when Aquinas says that "man and other rational creatures attain to their last end by knowing and loving God" (1a2ae 1, 8), he does so in the context of a chapter devoted largely to arguing in favour of Aristotle's thesis that man aims at happiness. It may therefore be a tenable interpretation to hold that whatever we gain through the intellect is merely instrumental, and that epistemic value, whatever that is, is not in any important way distinct from, say, the value of financial prudence.

Nonetheless, we must remember that the goods of the intellect and the goods of the will are not exactly commensurable. One aims at truth, the other at goodness. At times, both are involved in our intellectual endeavours (and, for that matter, our practical ones), but their conceptual independence is guaranteed by their separate ends. Those ends, as what grounds their formal objects, are what distinguish them as powers of the soul. It is therefore no stretch to think of them as having separate perfections, and therefore as seeking different sorts of values.

Deeper analysis seems possible, but would involve a very long detour into Aquinas' axiology and philosophical psychology. For my purposes, querying his epistemology is daunting enough—so I will leave the matter without further comment. Still, it is important to flag the potential reductionist reading, for it will have to be dealt with in what follows on Hume, as well as in chapter 6.

The outlines of Aquinas' answer to the Averroists (and, to some extent, to Plato) is made even clearer when we consider the ways in which our felicity is *not* achieved. Contemplation of God is our final end, but this is not the sort of epistemic relationship to God possessed by ordinary men, or by

[81] This ought not, of course, be understood as the Humean claim that understanding God is the ultimate end for all intellectual beings because it satisfies desire. Rather, desire in an orderly universe is a guide to perfection.

philosophers, or even by the faithful on Earth. How exactly it can be possessed is something Aquinas spends a great deal of time explaining, but is too large a subject to broach. It suffices for our purposes to say that God's grace perfects us in heaven such that we can see God's essence to some significant degree and without mediation, though we never understand God completely. Some concessions are granted to the Averroist. Faith "does not grasp the object to which it gives assent in the act of believing" (*SCG* III, ch. 40, [2]). It is "more like hearing than vision," for, "one who believes gives assent to things that are proposed to him by another person and which he himself does not see" (ch. 40, [4]). In this sense, demonstrative knowledge is superior. However, because it grasps less content (and less important content), faith surpasses knowledge (ch. 40, [1]). Thus, when Aquinas argues for all the things in which man's perfection does *not* consist, he treats faith last, while he treats more obviously bad candidates first. Knowledge gained through demonstration gets treated in between the two.

On this picture, the highest available epistemic good is the active contemplation of God, the center and cause of all existence. This is unavailable without grace, and seemingly unavailable on Earth. Here, the highest epistemic good lies in the imperfect knowledge of faith, for faith provides both truth of the highest sort and a low kind of understanding. But demonstrative certainty, though it involves understanding something, does not involve understanding God.

3.5: Hume:

Hume says relatively less about epistemic value, but what he does say stands in stark contrast to the ancient and medieval thinkers who precede him. What value there is in seeking knowledge comes from a few different sources; in brief Hume argues that the ultimate source of significance lay with the pleasure we take in exercising our own genius, that is, in figuring things out:

> …[T]he satisfaction, which we sometimes receive from the discovery of truth, proceeds not from it, merely as such, but only as endowed with certain qualities. The first and most considerable circumstance requisite to render truth agreeable, is the genius and capacity, which is employed in its invention and discovery. What is easy and obvious is never valued; and even what is in itself difficult, if we come to knowledge of it without difficulty, and without stretch of thought or judgment, is but little regarded. We love to trace the demonstrations of mathematicians, but should receive small entertainment from a person, who

> should barely inform us of the proportions of lines and angles, though we reposed the utmost confidence in both his judgment and veracity.[82]

Notice the stark contrast with Aquinas. For Aquinas, we come to the deepest truths by taking them on God's authority. And, as we have seen, we are only able to get that by our antecedent acquisition of faith through yet more second-order knowledge (the testimony of the martyrs, for example). For Hume, there could be nothing duller, and therefore nothing less worthwhile than taking something on authority or testimony.

It would be a mistake to impute from this any view on Hume's part that we ought to avoid second-order beliefs. Hume's concern here is to explain the philosophical psychology of curiosity as such, and while it may be that certain beliefs yield no pleasures in the vicinity of curiosity-satisfaction, nor allow the understanding to soar in glory, they might well serve other aims. A true second-order belief about where I can get my daily bread does not scratch any intellectual itch, but it is certainly important in satisfying other desires.

I draw attention to this because interpreting Hume to be somehow sidestepping the question of epistemic value as such is tempting, and, for all I know, correct. Any philosopher will recognize that there is a certain joy in pursuing and capturing the truth, and that this joy (likely) has some value just as joy. That pleasure varies according to our interest and our own role in the acquisition of the belief at hand. Why not think that Hume is simply interested in understanding *that* pleasure, one which has no strong connection to *epistemic* value as such. Such an account, after all, is entirely compatible with a view that there is *no* distinctively epistemic sort of value, for though epistemic endeavours sometimes bring with them pleasure, what is valuable in them just is the joyful feeling, and that joyful feeling would be just as valuable if obtained elsewhere. In any case, we might interpret the description in two ways:

1) Hume has a theory of epistemic value. Something is epistemically valuable just in case:

[82] Hume (1740), quoted in Whitcomb (2007: 47).

a. It comes immediately and directly from an intellectual endeavour of some sort.

b. It brings with it some kind of pleasure.

2) Hume has no particular theory of intellectual value. Curiosity is a remarkable disposition, and seems to be satisfied in particular circumstances, which he go on to enumerate.

Notice how the passage about the joylessness of randomly playing with conic sections, or the dullness of merely hearing the truth from a reputable source might seem to favor the value-realist interpretation of (1). After all, Hume is certainly claiming *some* value to be real, and that if this value is not identical to the pleasure of learning as such, then it is something other than what he means to define. But notice that it is also compatible with (2). Many epistemic activities bring pleasure, but only some bring it immediately. Those he means to discover, and we ought not to make any further inferences about his value epistemology.

The question of what Hume really meant is best left to specialists. For our purposes, it is enough to note that there seem to be at least two defensible interpretations in the tradition, and to note one feature they have in common: both are *internalist*. The epistemic prizes we seek are prizes because they make us happy in some way. Their worth comes not from their objects, but from us.

This is even more apparent later on, where he compares philosophical inquiry to hunting and gambling. I may have sufficient meat for my supper; yet I hunt. I may have surer financial prospects than a game of dice; yet I gamble. And yet, I would not hunt were there no game, and I would not gamble were there no chance of winning something. He explains why:

> But beside the action of the mind, which is the principal foundation of the pleasure, there is likewise requir'd a degree of success in the attainment of the end, or the discovery of that truth we examine. Upon this head I shall make a general remark, which may be useful on many occasions, *viz.* that where the mind pursues any end with passion; tho' that passion be not deriv'd originally from the end, but merely from the action and pursuit; yet by the natural course of the affections, we acquire a concern for the end itself, and are uneasy under any disappointment we meet with in the pursuit of it. This proceeds from the relation and parallel direction of the passions above-mention'd.

The life of the mind, then, is even more closely akin to a game in that we seek success because we get caught up in the joyful pursuit of it, and thereby come to desire that we win.

Hume's rather extreme explanation is unsatisfying. As Whitcomb points out, Hume doesn't think there is "*anything* beyond entertainment value that makes some truths epistemically more significant than others." It is implausible that the number of blades of grass on my lawn, to borrow an example from Rawls, is worth knowing in any sense beyond mere preference satisfaction, even to someone deeply curious about the matter. It might even be the case that the charitable reading of Hume is that he doesn't believe in non-derivative epistemic value at all. Rather, there is no epistemic value as such, except insofar as its acquisition or possession provides some kind of pleasure to its vessel.

Chapter 4: The Present Context:

4.1: Introduction:

Much of this dissertation is devoted to the axiology of epistemic content. Content axiology (CA) is largely undiscussed in the contemporary context. Most of the very small body of literature devoted to working out its details will be addressed in the next few chapters. Epistemic *state* axiology (SA), the branch of epistemology that deals with the value of various kinds of epistemic states—chiefly true belief, knowledge, and understanding—is much better canvassed. Because of the necessity for *some* connection to present debates, I situate the CA question in the SA dialectic. My arguments are overwhelmingly of the negative sort. I will attempt to show that contemporary philosophy has failed to demonstrate many implications between the two fields. Null results are always a disappointment for new projects. But whether or not it helps start the debate I wish philosophers would have over CA, those are the results I have found.

This chapter is in four parts. The first deals with the relationship between CA and more popular debates about veritism and monism. Veritism, roughly speaking, is the view that truth is of fundamental epistemic value. Monism, equally roughly, is the view that there is just one source of fundamental epistemic value. The two views are usually conjoined, and I will use "veritism" to mean the view that *only* truth is of fundamental epistemic value unless otherwise stated. Veritism is sometimes criticized on the ground that it cannot make sense of our intuitions that truths differ in their value. If only truth matters, the argument goes, then we have no basis for these intuitions. But this is not so, for, as we will see, veritism is compatible with truths differing in significance. I conclude somewhat disappointingly: for all that has been said in the main contemporary discourse surrounding CA, it is not the case that much follows regarding SA.

The second and the third parts concern a popular argument against veritism. DePaul (2001) argues that veritism must be false because, first, sets with equal numbers of truths sometimes have different values, and, second, that veritism implies that those sets have equal value. As such, veritism is false. The second section of this chapter is a short argument in favor of DePaul's first premise. At

the lowest level of resolution, the argument is that the union of sets of truths is sometimes more valuable than the constituent parts of the union. The third part argues that DePaul's more controversial second premise fails. *Pace* DePaul, sophisticated construals of veritism do not imply that all truths are of equal value. Instead, we can understand truth and significance as making up a kind of organic whole.

The fourth part reviews what little literature there is on how to conceptualize and formalize CA. I do not mean that it goes into any detail on the wellsprings of CA. Rather it offers a scheme for expressing quantitatively how some true, significant beliefs might be more or less valuable than others. Finally, I argue against Whitcomb's view that this implies very much about veritism specifically or SA generally.

4.2: Content Axiology and Veritism:

I will begin with some surface-level remarks on why veritism is plausible. Veritism holds that only true beliefs are of basic epistemic value.[83] Vertism is the most popular version of what Pritchard (2007; 2009; 2019) calls *epistemic value monism*. If all epistemic value lies in getting to the truth, then the only state that *fundamentally* matters is whatever state gets us to the truth—in this case, mere belief. Pritchard (2016a) himself distinguishes between believing and grasping, but grants that the distinction is slight enough to be set aside (2016b). It is more-or-less dialectically safe to stipulate that all gettings-to-the-truth, be they cases of knowledge, understanding, or whatever else, necessarily involve *believing*.[84] Believing the truth, on the most intuitive version of monism, is what we aim for on veritism, and anything else (say, having good reasons) is auxiliary.

If we assume that there is only one (bedrock) valuable epistemic state, then true belief is a pretty good candidate. But why think there is only one? The most obvious attraction is parsimony. Consequentialists usually admit that their views have unpalatable entailments (e.g., that it is sometimes obligatory to torture the innocent), but hasten to point out that any system that

[83] Sosa (2001), Goldman (2002), and Alston (2005) all arrived at this view at roughly the same time.
[84] See Wilkenfeld (2017) for an opposing view. See Hills (2016) for a nuanced, moderate view.

accommodates *all* our first-order moral intuitions is likely to grow implausibly large and baroque. Occam's razor is as foundational a methodological principle as any in philosophy.

More subtly, monism seems to link the epistemic with everyday practices. Most people don't use phrases like "epistemic value" of course. As Pritchard points out, the term "epistemology" didn't even appear until the middle of the 19th century. However, people everywhere do understand the difference between getting to the truth and failing to do so. Pritchard offers the following example and brief analysis:

> The folk would surely recognize… the important difference between believing a person charged with murder is guilty because of the overwhelming evidence available to support that verdict, as opposed to believing that he is guilty because of some feature of the defendant (his skin colour, say) which one is prejudiced against. Thus, if the epistemic is concerned with the grasp of the truth, then we can plausibly contend that it picks out a certain kind of evaluative domain which is rooted in our everyday practices, such that it isn't merely a term of art (2016a: 3).

The idea is that epistemic axiology (or, perhaps, epistemology generally) is consonant with ordinary human practices. Other kinds of axiology seem to have visible roots in human practices. A painting can be valuable aesthetically, intellectually, historically, or economically in distinct ways that everyone can appreciate. Its being beautiful *or* ugly is compatible with its being interesting *or* boring, and with its being original *or* derivative, and with its being a bargain *or* overpriced. But its being ugly does not make it boring, and its being original does not make it a sound investment. Thus, the aesthete need not worry if her calling a picture valuable comes into (surface level) conflict with what an art appraiser says about it. Everyone knows as much, and this should reassure the aesthete somewhat, for her work is geared towards sorting out a kind of value that we recognize.

The pluralist will have more trouble justifying her endeavors. As a toy example, take the maximal pluralist, someone who thinks that all of the following epistemic categories have scales of value proper to themselves (i.e., irreducible to those of their neighbors): Belief, knowledge, understanding, justification, rationality, certainty, truth, agency, explanatory power, being Gettier-defeating, resulting-from-a-reliable-process, and social-epistemology-appropriateness. Such a person will have a great deal of difficulty explaining what epistemic value is! For similar reasons, I expect

such a person will not really have a carved-at-the-joints account of the practice of epistemology itself. Why bother with it at all?[85]

Of course, none of this is dispositive. Disciples of W. D. Ross in ethics may bristle at the charge above, averting to the world's stubborn complexity and the hubris of oversimplification. But it is far from clear that this is a big worry for veritists. Ross, after all, thought that these pro-tanto goods were, as it were, united in the all-things-considered analysis. That, in the end, there is a right thing to do, and that though such rightness is a river with many ducts and tributaries, it is the business of ethics simply and straightforwardly. I'm not sure the maximal pluralist about epistemic value can say the same. She has no final epistemic good to unite the above. Little more can be said about this worry at this cursory level.

In order to see why any of this matters for CA, a longish inquiry into the relationship between truth and value is warranted. At the beginning of this chapter I mentioned that contemporary debates about SA intersect with the largely ignored matter of CA. DePaul, an SA theorist who opposes veritism, argues that, because all true beliefs must be of the same value if all that matters is truth, and because not all true beliefs are of equal value, then veritism is false. This has costs alluded to at the beginning. For one, veritism is parsimonious, and any non-reductivist denial of it will end up multiplying and separating lots of different kinds of epistemic value. Second, this kind of baroque axiology threatens to destabilize the discipline. If there is no basic epistemic value to unite epistemology, then we might have well-founded worries about the discipline as a whole. Analogously, Utilitarians and Kantians, however subtle their views might be, generally take pride in the simplicity of the foundations of their views—that is to say, the unity of the thing (pleasure, adherence to rational duty) that makes all good actions good. As such, investigating the worry is no idle exercise.

4.3: Atomic Beliefs and their Value:

[85] Note that this is not so much a consideration in favor of veritism specifically as it is in favor of monism generally. A "knowledge first" view (what Whitcomb (2007) calls "epistimism") will work just as well for all I have said.

Our investigation begins by looking at the way true beliefs relate to their parts. If I believe *p*, where *p* is the explicit conjunction of *q* and *r*, then I normally believe *q* and *r*. This may not be exceptionless, but it seems to be true in normal cases. Plausibly, whatever epistemic value I possess from believing *p* is equal to the sum of the epistemic value of believing *q* and the epistemic value of believing *r*.

If this is right, and I will argue that it is not, an interesting philosophical project presents itself. We can conceive of all the true propositions as existing in various relations of conjunction, disjunction, entailment, and so on. Those relations operate over what we can call "atomic" propositions, or those fundamental propositions being conjoined or entailed and so on. The sort of project I have in mind is a rough analogue to a kind of mereological reductionism. We can conceive of the world as composed of only simples, where what truth we speak about collections of simples is reducible to the constituent truths about constituent simples. So when I know that all protons have positive charge, I am in possession of exactly as much epistemic value, so far as protons are concerned, as someone who knows that each proton has a positive charge. When I believe something true about simples-composed-elephant-wise, I am in possession of the epistemic value inherent in those simples. But I am not in possession of anything more. The atomic truths concerning atoms are of fundamental value; everything else is derivative.

More boldly still: when I believe a complex scientific theory, I believe a huge number of truths, and I am in possession of a trove of epistemic value equal to the constituent truths of that theory. We can conceive of towers of implications and conjunctions, where many atomic truths are purchased at once by our buying a truth that ranges over them. A theory along these lines makes sense of an intuition I suspect many share: that the mark of a deep, valuable truth is that it brings with it large numbers of other truths. The questions of physics, cosmology, philosophy and theology—those we treat with the greatest respect—are questions about all of reality, and their answers cast wide shadows over everything.

The challenges such a theory faces are immense. How do we go about counting this indefinite (infinite, perhaps) pile of truths? Do we believe an infinite number of things in the first place? How

exactly does entailment work here, when we so often do not realize the truths entailed by the truths we know? Or conjunction? On the other hand, the line of reasoning is not obviously flawed from the outset. It does seem that the value of composite truths is at least partly a function of the value of their components. The project fits the reductive impulse that so often clears away philosophical cobwebs; perhaps it has value.

In any event, the reader may be wondering what any of this has to do with the intersection of CA and SA. It pertains because of a salient critique of veritism. Veritists, recall, hold that all non-derivative epistemic value lies in the possession of truth. DePaul (2001, 173-74) argues that veritism is false because two sets of truths with equal numbers of members need not be of identical value. But since veritists hold that only truths are valuable, then they have no way of differentiating between the value of two sets with equal numbers of members. Therefore, veritism is false. In the following section I will lay out the reasons for rejecting the second premise of the argument. There are not, I argue, any implications to be drawn for veritism from content axiology, at least not in this way.

If I am mistaken in the next section, then the veritist can still reject the first premise. For, if the atomistic project outlined above can gain steam, it may be that sets of truths with equal numbers of members are equally valuable. In this section, I argue that *this* attempt to rescue veritism is a non-starter. The first premise of DePaul's argument—the premise that denies that equal sets of truths have equal value—is correct, and concerned veritists should focus on the second premise.

To see why, we examine Timothy Perrine's (2021: 12115) tentative defense of the premise.[86] He labels DePaul's objection the "uniformity" objection, since the objection takes issue with veritism's presumed labeling of all truths as equally valuable. Now, Perrine's preferred veritism, "Sophisticated Veritism," divides truths into "basic" and "non-basic" categories along the lines suggested above. Basic truths compose non-basic truths through operations like conjunction. Since

[86] Perrine does not fully endorse the view I am ascribing to him here. Rather, he seems to think it has a better chance of being true than it gets credit for.

non-basic truths are sometimes more valuable than basic ones by virtue of basic ones being proper parts of non-basic ones, the first premise of DePaul's argument can be rejected.

We can amend DePaul's argument to preserve the force of the objection. Since, for Perrine, all epistemic value derives from basic beliefs, the reformulated first premise is that sets of beliefs that contain the same number of basic beliefs are not always of equal value. Above, I laid out the barest bones of a program for biting the bullet. If all grand beliefs turn out to be collections of individually insignificant basic beliefs, then perhaps there is no problem with arguing that all basic beliefs are of equal value. Fortunately for us, we can avoid this rabbit hole. For even if this view turns out true, the value of "large" summative beliefs ranging over many atomic beliefs is not equal to the sum of their atomic parts.

Occasionally, one hears a philosopher joke about how strange it is that he happens to have only true beliefs. "How bizarre," she will remark, "that every single thing that I believe is right!" But the reason such jokes are amusing is that of course the speaker knows fully well that many of her beliefs are false. The paradox in the vicinity is usually called the *Preface Paradox*, for one frequently sees book prefaces containing apologies for inaccuracies contained therein. There remains some contention about how to resolve the paradox, but whatever solution ends up being correct must allow us to believe that some of our own beliefs are false, or at least to refrain from believing that we hold only true beliefs. Anything to the contrary would be a staggering blow to the certainty we all acquire in our own fallibility (or should acquire, in any case).

Good sense demands that we refrain from forming large summative beliefs some of the time. But not all the time. Sometimes, large summative beliefs are rational. Examples abound, many concerning authority. If I discover Nico is a compulsive liar about his relationships, I can rationally believe what I already suspected—that each thing he told me about his relationship is false. When summative beliefs are rational, and they turn out true, they are more valuable than the sum of their basic parts. Suppose I believe the proposition A&B&C&D. On Perrine's view, the total final value of my belief is equivalent to my believing that A, and my believing that B, and so on. But this is false. I might believe in each individually, but still not take a bet that their summation is true, and given a few

obvious assumptions (e.g. that I lack sufficient certainty in the unconjoined propositions) I might be perfectly rational to do so. Admittedly, success in betting is no obvious measure of epistemic value. But I do not think I need any such argument since it is independently intuitively convincing, and widely held, that believing each of a set does not imply that we do, nor that we should, believe the whole set. Alexander Pruss (2019) argues for God's existence by appealing to the summative belief that each and every skeptical scenario that philosophers have dreamt up fails to obtain. So, it is false that we are brains in vats; it is false that we are Boltzmann brains; it is false that demons deceive us about most things; it is false that we are simulations; it is false that everyone else is a P-zombie, and so on. Now, on atheism, there may be good reason to buy each claim. We know we are not brains in vats because science has not advanced sufficiently far to envat brains; we know we are not Boltzmann brains because evolution is needed for brain function; we know that we are not deceived by demons because there are no demons, and so on. The argument, very roughly, is that theists have reason to think that *all* of these scenarios fail to obtain, while atheists have (at best) only reason to think that *each* fails to obtain. Whether the argument is sound is beyond me, but it seems correct in its form.

A credence approach to belief will not help. Suppose we dispose with a trivalent model of belief (or quadrivalent, depending on how successful the first chapter of this thesis is), and simply analyse all doxastic stances in terms of credences. On this view, if the angel Gabriel tells me that all my beliefs are true, my credence in each rises to 1. As such, any new belief in their conjunction adds nothing. But this defense has dim prospects. Arkady and Belinda are mathematicians with excellent knowledge of a particular proof. They know with perfect credence *a*, and that *a implies b*, and that *b implies c,* and that *c*. Suppose Arkady understands entailment well. He believes the conjunction of *a*, *a implies b, b implies c,* and *c*. Belinda, however, is irrationally underconfident. Despite knowing each step with certainty, she also believes that it is never rational to believe large sets of propositions with certainty. When Gabriel reveals to both of them that the conjunction is true, Arkady learns nothing new—while Belinda learns something. As such, trying to repair the argument by importing a very controversial credence-based account of belief is fruitless.

The project of atomizing belief is an interesting one, and may be very appealing to certain kinds of very hard-nosed reductionists. But it does not tell us very much about the debate about CA and veritism, because it relies on an implausible understanding of conjunction.

There are other criticisms of DePaul's first premise. Treanor (2014) argues that the first premise is false because there are not two sets with an equal number of members. If one set of true sentences contains only a member holding that "grain of sand x is so-and-so-many millimeters from grain of sand y" and another set contains only a member holding that "stars are formed when massive clouds of molecular hydrogen collapse due to gravity," do these sets, if believed contain an equal number of true beliefs? They certainly contain an equal number of true sentences. But consider that a natural way to read the first set is as "containing" the following distinct elements:

1) x is a grain of sand and

2) y is a grain of sand and

3) x is so-and-so-many millimeters away from y. See Treanor (2014, 556), quoted in Hu 2017, 259).[87]

One might even think that it contains further truths, like "x exists" and "y exists." The problem is familiar from the previous section—counting truths is very difficult. This is since, argues Treanor, we have no reason to think that there are two sets of similar size.

There is a hitherto unmentioned distinction between basic and intrinsic value. Something is intrinsically valuable just in case it is valuable and not merely as a means to something else valuable. Something is of basic value just in case it is valuable and not merely in virtue of a proper part being valuable. Following Hu, we can abbreviate "basic intrinsic epistemic value" as "basic epistemic value" and "intrinsic epistemic value" as "epistemic value." Basic value can drop out because

[87] Treanor's argument is explicitly endorsed in Prichard (2014).

everyone knows that it's what we're looking for, and few think the distinction is misunderstood in a relevant way.[88]

Here is Hu's summary of DePaul's argument, the gist of which we are already familiar:

1) There are two sets that contain an equal number of true beliefs but differ in epistemic value.

2) If veritism is true, then any two sets that contain an equal number of true beliefs have equal epistemic value.

3) Therefore, veritism is false (3).

DePaul defends the second premise by asking us to compare some known empirical theory to a set of beliefs containing random, easy, and unrelated examples of arithmetic sums. Plainly, he says, the first set is more valuable, but it is perfectly possible that the two sets contain the same number of true beliefs (2001, 73). His defense of the first premise is light, and need augmentation.

Hu offers two responses to Treanor's objection in defense of DePaul's first premise. We will look at them both before turning to Hu's critique of DePaul's second premise. "[A] person who believes that *grain of sand x is so-and-so-many millimeters from grain of sand y* does not have to believe all the truths behind the door. He might just believe the… truths Treanor lists above and no other truths. After all, how many truths a person actually believes is a psychological matter, not a logical issue" (5). Treanor is engaged in the important project of determining the relationship of true propositions to each other. What Hu is drawing our attention to is the fact that truths need to make their way into our minds in order to count as valuable possessions. Entailments that we do not grasp our not valuable for us in this way.

I am not sure how far this response goes. The trouble lies with the difficulty in counting beliefs "psychologically." Take a simple mathematical case: All multi-digit positive natural numbers are greater than four. We all believe that ten is greater than four, as is eleven, and so on. Do we

[88] See Perrine (2018) for an opposing view.

believe that some large, arbitrary number (4,849,202) is greater than four, supposing we have never entertained the question? My sense is that we do. Now it is another question as to whether that belief is "contained" in our other beliefs. If beliefs are sufficiently dispositional in nature, and many philosophers think that they are, then it might be that if we have sufficient understanding of mathematics to grasp principles like "double digit positive integers are larger than single digit positive integers" then our believing those principles might contain an indefinite number of beliefs about specific inequalities. As we have said previously, the relationship between implication and "containing" is tricky. But it is not obviously wrong to think that (roughly) comprehension of x plus belief in x plus a certain level of intelligence just is to have a belief in x that "contains" some of x's entailments. If this is right, then perhaps Treanor's argument can be rescued, for it would remain unclear whether we can find two identically large sets of beliefs.

Fortunately, Hu has a backup response. Finding equally sized sets might be difficult, but it is surely possible to imagine two sets of clearly different magnitude, where the smaller set is more valuable. Hu's example is a good one. Compare the set of beliefs containing every truth about how many commas are on each page of every book published since 1990 to the set containing just the view that oxygen is necessary for combustion (6). Plausibly, the second set contains many truths ("oxygen is a gas", "oxygen exists", "combustion exists", etc.), it is far less plausible that it contains billions. As such, Treanor's response fails.

4.4: Veritism and Truths of Disparate Significance:

DePaul's reductio is the most common criticism of veritism. (There are, of course, others, but those tend to be far more complex and require more commitments.) In this section I will look at two more responses to DePaul's reductio. One is Ahlmstrom-Vij and Grimm's; another is Hu's own. Little I have to say has not been already said by Hu, with whom I agree wholesale. Since I have little original commentary to add, I will be brief.

Ahlmstrom-Vij and Grimm take a different, and *somewhat* more fruitful approach. Unfortunately it suffers both from the difficulties inherent in Humeanism and further conceptual error.

The problems with Humeanism I have already discussed some, though far more will be said later. The other error will become clear in a moment. First, the argument: Ahlstrom-Vij and Grimm think, correctly, that beliefs vary in value. But how can we say this, if truth is the only thing of basic epistemic value? "In short," the answer, "by defining the epistemically valuable in terms of that which is valuable in the way of inquiry and identifying inquiry with the inquisitive practices concerned with… answering questions, and then characterizing the significant as that which pertains to questions that we want answered" (2013, 332). They have avoided the worry that might plague what we might call Platonists about CA (how can one truth "contain" more truth than another?), but only by committing to the view that significance just is whatever we want answered. But that aside, they take themselves to have given an account according to which there is only one basic epistemic value:

> Granted, significance is a property of (some) true beliefs. … [But] it is not a property that true beliefs have independently of our conceptions of what makes for a worthwhile inquiry. More specifically, significance measures the degree of epistemic value as a function of the extent to which the relevant true beliefs speak to inquiries that we deem worthwhile… The mistake of the objection under consideration is the assumption that a significant true belief has two properties of epistemic value, i.e., truth and significance, as opposed to one property of epistemic value, in great quantity (2013, 334), quoted in Hu (2017, 263).

As Hu points out, there is an implicit assumption that "a thing that only determines the degree of the epistemic value of truth [namely, significance,] is not itself of basic epistemic value. Take two beliefs of differing value. "Ahlstrom-Vij and Grimm think that [one] is more valuable than [the other] because [one] is more significant… But both [beliefs] have the property of being true. So Ahlstrom-Vij and Grimm have to admit that the epistemic value of [each] cannot be fully explained in terms of the epistemic value of its property of being true" (9). Since Ahlstrom-Vij and Grimm hold that truth is the only thing of basic epistemic value, we should expect the contrary.

Hu's proposed solution requires us to abandon Ahlstrom-Vij and Grimm's particular flavor of veritism. Their view is that veritism holds that the *property of being true* is the only basically valuable epistemic good. Hu's is subtly different. For him, "only true beliefs are of basic epistemic value" (262). This, in fact, was what veritism was always supposed to be. Not only should the veritist deny that the property of being true and the property of being significant are valuable. Rather,

> she should maintain that none of the properties of a true belief can singly make the true belief epistemically valuable. Rather, it is the property of being true and the property X that co-make some true beliefs some true beliefs epistemically valuable, where X could be the property of being significant or some other properties of those true beliefs (265).

For Hu, the property of being significant and the property of being significant (and possibly other properties, for all we have said) co-make beliefs valuable. This kind of "organic whole" is what the veritist should be after, rather than either alone. This will not face the kind of difficulties that the previous view faced, for it allows the veritist to say that all and only significant truths are of basic final value, but without inconsistency. How many beliefs end up being significant is a separate question, but it has the minor virtue of being compatible with views like Kvanvig's (2007), according to which every truth is of some (frequently minimal) value, as well as views like Goldman's (2002), according to which some truths are without value.

Hu's view is compatible with a Platonism, Thomism, and Humeanism about CA. One way to bring this into focus is to look at his answer to a potential objection. Here is the objection, in his own words:

1) Veritism states that only true beliefs are of basic epistemic value.
2) A basic epistemic value of a thing is, by definition, an intrinsic value from an epistemic point of view.
3) The intrinsic value of a thing from a certain point of view solely depends on the intrinsic properties of that thing relative to that point of view.
4) There is only one intrinsic property of a true belief relevant to an epistemic point of view: the property of being true.
5) So according to veritism, the basic epistemic value of a true belief solely depends on its property of being true. (261)

The third premise implies that any objects with the same intrinsic properties have the same intrinsic value. This is highly controversial. Forgeries, Hu points out, seem to be less intrinsically valuable than their originals in virtue of *inter alia* their not being original. But forgeries have the same intrinsic properties, so (3) is dubious, though by no means obviously false. Furthermore, (4) is suspect. If we assume a popular "correspondence" view of truth, then the property of "being true" is *not an intrinsic property of true belief*. Hu explains:

> The property of being true, by definition, is the property of corresponding to reality. So a true belief is a belief that has the property of corresponding to reality. The property of corresponding to reality is a relational property shared by all beliefs that are true. A relational property, by definition, is not an intrinsic property. So the property of corresponding to reality is not an intrinsic property of a true belief. That is, the property of being true is not an intrinsic property of a true belief (262).

Now, it may be that the correspondence theory of truth is correct. I am happy to assume that it is, in part because of my own overwhelming intuitions on the matter—but in part because many of the thinkers involved in the debate, as Hu points out, agree.[89]

Arguing for or against veritism is not my primary aim here. But I think there is another valuable lesson nearby. Above, we saw an argument for veritism on the basis of its parsimony. The unity—and therefore utility—of one of the main branches of philosophy depends on the unity of the epistemically valuable. Recall the importance, for Prichard, of such unity. Is this unity preserved in Hu's account? Yes and no.

Hu preserves veritism. He sets out to show that a popular argument against the thesis that "only true beliefs are of basic intrinsic value" is unsound. As far as I am concerned, he has done this. But it was not without a cost. On his picture, a belief's being true (that is, its property of corresponding to reality) is not the only element making a belief valuable. Rather, it co-makes beliefs valuable along with something else (more precisely, a further relational property), namely, the property of being significant. One wonders how costly this is for the veritist. The answer, I think, depends on whether she favors a world-based Platonic or Thomistic veritism or an agent-based Humean determinism. Suppose she falls in the first camp. Epistemic value, then, depends on a whether the belief in questions not just *conforms* to reality (i.e. is true) but whether said belief adequately *maps* or *captures* enough of reality. Phone-book beliefs can be true, but they *capture* such a tiny sliver of reality that they fail to be valuable. The business of inquiry, then, is still relatively united. It is the task of the inquirer to figure out, in a substantive way, what the world is like. One must seek to get at the whole picture, as it were—or at least enough of the picture, in order to qualify

[89] Goldman and Alston, among others, endorse this view.

as gaining epistemic value. Crucially, however, there is a real kind of organic unity here. Platonist and Thomist veritists are still fairly parsimonious, for they value figuring out how the architecture of reality works, even if they think some bits of it more-or-less extraneous to that goal. Admittedly, the view under examination is not quite as simple as the "truth-only" version DePaul attacks. But it has the view of being immune to his criticisms while retaining an intuitively unified goal. It can be made sensitive to what Sider (2011) calls the structure of reality—that is, sensitive to questions that cut reality at the joints. It can be made sensitive to the view of a unified, orderly, centered reality that seems to underlie the views of so many philosophers, and which we will return to in detail next chapter.

But notice that the Humean veritist cannot be quite so frugal. Her version of Hu's view may, of course, still maintains the coherence of veritism. After all, she may think that *our* finding questions interesting or important might confer the property of significance that co-makes some organic unities valuable, and do so without ever denying that all and only significant, true beliefs are of basic epistemic value. She does not stick her toe outside the technical, definitional confines of the view. But being satisfied with this is to over-value the careful, philosopher-nuanced definition. For, she cannot gesture to the unity of the true and the significant, wherever she finds it, out there in the world—already, as it were, bound up with each other. Rather, the "X" that, along with correspondence to reality, co-makes beliefs basically valuable, lies somewhere in *her*. For this reason, she will have a much harder time staying on the right side of Occam's razor in her debates with anti-veritists.

4.5: Content-Significance and States:

Last chapter we saw that Plato thought the best kinds of epistemic states, those that amounted to the loftiest kinds of knowledge, were knowledge of the highest matters. Aquinas' view departed radically. For him, the most valuable truths could not be had with deep understanding, though they could (perhaps) be had through what we today would call knowledge. There remain a (very) few contemporary views on the subject worth investigating. Goldman (1999: 2002) and Whitcomb (2007) have competing views, labelled "veritism" and "epistemism" respectively, that have something to say about epistemic value and content. (Veritism we already saw in the last section, where I assert that it

best fits a Platonic model, not a Humean one.) Goldman thinks we can make sense of the value of content in terms of truth that we care about. Whitcomb argues that the value of content is best understood in a framework where epistemic value is a primarily a function of proximity to knowledge. I argue that Goldman's view is mistaken, and that Whitcomb's view is really two views that have little to do with each other, only one of which is helpful.

Goldman's veritism is a broadly Humean one concerned with the value of true beliefs. This is not the only sort of valuable thing for him, nor is it the only sort of valuable thing in epistemology. His veritism concerns *only* the way the beliefs are valuable or disvaluable vis-à-vis truth. Now, Goldman is sensitive to the problem of the *epistemic* absence of value in insignificant truths.[90] Maximizing true belief is important, but not all truths are equal, and we want to make sure we get the right ones. His solution is bivalent. Either we are interested in a given matter, and thereby might gain something of Veritistic value, or we are not, and are dealing with something veritistically inevaluable. It is not the case that my knowing some totally inconsequential, dull, tiny bit of trivia is either valuable or disvaluable; rather, it fails to achieve a threshold of interest. I will follow Whitcomb (see his 2007, 139) in interpreting Goldman's "interest" to be functionally synonymous with "significant," though of course we may be wrong about this. By "significant" I mean of distinctively epistemic non-instrumental value, rather than e.g. practical value. By "interest" I mean the sort of thing about which agents are concerned.[91] The view is Humean: whatever interests us is what is significant. The interpretive move is warranted. Should Goldman think there are two senses in which something might be veritistically valuable, he would likely have said so. Furthermore, it seems odd to deal with the problem of trivial truths by appealing to our interest if it were *also* the case that they were too insignificant to qualify as the sort of the thing that can be valuable regardless of our interest. As such, I'll use "interest" and "significance" interchangeably when discussing his work. Even if I am wrong, then I am still addressing a theory with enough *prima facie* credibility to be worth addressing. So,

[90] Other dimensions of value, which may or may not explain how and whether insignificant truths have value, are not what concern Goldman here.
[91] I believe the view is compatible with a normative view of interest ("what she would care about if she had any sense") and a descriptive one (what she cares about). The reader can insert whichever she thinks is more plausible.

then, if a proposition is insignificant, then it is neither disvaluable nor valuable. If it is significant and true and believed, I have achieved a veritistic utility "V-value" of 1. Should I believe something false, my V-value is 0, ad if I withhold judgment, the V-value is .5 (1999: 89).[92]

DePaul (2004: 94-6) and Whitcomb (2007: 141-2) both point out the strangeness in taking insignificant propositions to be veritistically inevaluable. Suppose one person believes truly that *p*, something insignificant, is the case; another believes falsely that it is not. Surely *from the point of view of the truth*, the former is doing better. This, as Whitcomb puts it, is a "paradigm case of epistemic superiority." I think Whitcomb and DePaul's position is correct, but suspect it is not persuasive. Trivialities are inevaluable because they do not matter; doing better on something that does not matter does not make one better off where it matters. To put it in the language of algebra, there is no such thing as "negative zero."

Whitcomb offers another argument (144). One can think of another sort of comparison. Suppose that Anatoly believes something *true* and *significant*, while Beth believes something *false* and *insignificant*. Had Beth believed something false and significant instead, Goldman would be happy to grant that Anatoly is doing better. Yet, in virtue of the triviality of her belief, where triviality is presumably "an epistemic *badmaker*," we cannot say Anatoly is doing any better. But this argument too is unlikely to persuade, for several reasons. First, it is not clear how this comparison is meant to work exactly. We are not told what Beth thinks about the significant proposition. If she does not believe it, then clearly Anatoly is doing better overall, even by Whitcomb's lights. If she does, then we are back to where we were in the previous paragraph: trying to convince Goldman, who does not believe that insignificant propositions are V-valuable, that being right about an insignificant proposition makes us epistemically better off. Furthermore, it is unclear that insignificance in error is a badmaker at all. I would much rather one of my false views was insignificant than significant.

[92] A reader asked whether propositions like "this is an insignificant proposition" are themselves insignificant. Goldman does not say, but I imagine the answer is contingent on the rest of the agent's cognitive situation. If I am very interested in selecting an example of a trivial belief for a philosophy paper, the meta-belief about the trivial belief may be valuable. Other meta-beliefs about insignificant matters may not have that dignity.

Another of Whitcomb's attempts fares little better: "A person who *believes all* of the true insignificant propositions is clearly in some way better off epistemically than an otherwise identical person who *disbelieves all* of the true, insignificant propositions. Furthermore, if an epistemic difference between these people exists, then that difference must entail a veritistic difference, because the only differences between these people concern whether their beliefs are true" (146). This argument is a combination of a misplaced intuition and insistence. The insistence is that even if we can see how one worthless belief counts for nothing, surely two or ten or all of them count for something. Again, while Whitcomb's position seems correct to me, it is not likely to convince. The second element concerns inferences. Suppose that knowing the location of any single carbon atom is worthless. It would still be the case that if one knew the location of all of them, one would be in a very good position to infer all sorts of useful information. But, of course, that sort of inference is an inference to a significant proposition one does not *now* know the truth of. Goldman's view is not, I take it, that one can never infer something significant from something insignificant; rather, it is that only the significant truths are V-valuable. Insignificant propositions may be useful as means to other propositions, but this is obvious. Consider a madman who will grant you perfect knowledge of a unified theory of physics if you can sincerely tell him how many blades of grass are in John Rawls' garden. Of course such knowledge is significant instrumentally. Or, should Goldman want to maintain some rule about how any proposition the possession of which leads one to significant truth is itself significant, then he would be within his rights to say that once I know enough truths about carbon atoms I start knowing something significant. Whichever is the case, Goldman's defenders have recourse to fall back upon.

I don't think a well-worked out account of epistemic significance needs to commit to the view that there are no completely insignificant propositions—or, indeed to its opposite. What it plainly does need is some way to deal with degrees of significance. No sane person thinks that significance is bivalent in the sense that it cannot come in degrees. Goldman himself implies nothing of the sort; he merely thinks that insofar as significance relates to *veritistic* value, that value having to do solely with the value of the truth of a belief, that it can render some beliefs inevaluable. But, for reasons DePaul

and Whitcomb point out, this is a somewhat peculiar conceptual juxtaposition. There may be truths not worth having, but from the point of view of being true, they seem to have as much "value" as any other true proposition.

Here's a clearer way of saying what I mean: Suppose that Goldman accepts the obvious thesis that some truths are more significant than others. Now, it is either the case that they have higher V-values than less significant (but still significant) truths, or they do not. If they do not, then we should understand V-values to gauge how much truth has been mined from a given proposition. Suppose they are not significance-sensitive, if I believe a true proposition with a reasonably high credence, I might have gleaned a V-value of .9, say, independently of how important the proposition is—*unless the significance is 0*, in which case the V-value does not exist. But this last clause seems unmotivated, for degrees of significance exist, and they presumably have a multiplicative effect on value[93]. It does solve the problem of worthless truths (if there is such a problem), but this is just as easily solved by setting the significance multiplier to 0, as it were. Just as the falsehood of a proposition renders my emphatically believing it disvaluable, so does its insignificance.

This is, I think, the best reason for rejecting Goldman's view. The problem is not that it is obviously untenable, or that it leads to any contradictions, or that it cannot deal with tricky cases. The problem is that it doesn't accommodate one plain truth. Significance clearly comes in degrees, so the value of (true, believed) propositions for their believer comes in commensurate degrees. It is simpler to understand significance as a multiplier of value for all propositions, where that multiplier might (for all I have said) be 0, rather than to have it make a difference only for insufficiently significant ones.

Whitcomb's final argument is slightly different in character, but the same in spirit. He assumes Goldman's view, then argues:

> Now, consider propositions that have arbitrarily low positive degrees of significance. Given [Goldman's commitments], people *are* penalized for lacking beliefs in these propositions of epsilon significance, despite their not being penalized for lacking beliefs in propositions of zero significance. But if it is a genuine problem to veritistically penalize people for lacking

[93] Consider how strange it would be if they had additive value.

> beliefs in true propositions of zero significance, then surely it is also a problem to veritistically penalize people for lacking beliefs in true propositions of epsilon significance. Indeed, the problem with true propositions of epsilon significance is the very same problem as the problem with true propositions of zero significance. In both cases, there seems to be something wrong with penalizing people for lacking true beliefs of such little significance. In the epsilon case, the "little significance" just happens to be a bit larger… And how could a mere epsilon of difference in significance make penalization not a problem in some cases, whereas it is in others? (156).

Whitcomb sees no good reason to let significance make a difference in only one case—the case where it reaches a sufficient level to count as evaluable—but nowhere else. I add merely that Goldman's theory is sparser if we conceive of significance as having a multiplicative effect everywhere, rather than just at the lowest margin.[94]

Indeed, Whitcomb's own view, *epistemism*, does just that. Epistemism is a thorough and comprehensive view, and not one that can be summarized quickly. At its crux is the thesis that epistemic goods, at bottom, are not a matter of true belief, but of knowledge. So, for example, justified false belief is more valuable than unjustified false belief, and not merely because pursuing justification may e.g. make it more likely that true belief can be achieved in the future. Justification, because it is knowledge-constitutive, is an epistemic end in itself. This, then, is its first and most central thesis (1) "Whenever one belief instantiates more of the knowledge-constitutive goods than another, the former is better than the latter, epistemically as an end, *ceteris paribus*" (245). Epistemism also contains other theses: (2) Withholding is a neutral end, (3) Virtues improve epistemic ends, (4) Degrees (of knowledge-constitutive goods) improve epistemic ends, (5) Epistemic states can be epistemic means, and, most importantly for us, (6) Significance amplifies epistemic value.

This sixth principle is spelled out in the following way:

> Suppose that A and B are token beliefs such that the proposition at which A aims is more significant (for the person who holds A) than is the proposition at which B aims (for the person who holds B). Then,

[94] Of course, Goldman's numerical scheme will have to be jettisoned. On his view, fully held false beliefs (that is, with a credence of 1) have a V-value of 0, so significance could not act as a multiplier that would make any kind of difference, since anything multiplied by 0 is 0. It would be better to set a false belief at -1, and set total agnosticism at 0. I note this in passing only, for I do not have more to add to the mathematical formalization than this.

- If principles 1-5 render A of equal epistemic standing (as an end) as B and that standing is *positive*, then A is epistemically better as an end than B.
- If principles 1-5 render A of equal epistemic standing (as an end) as B and that standing is *negative*, then A is epistemically worse as an end than B (245-46).

The principle is a more thorough articulation of the view that significance enhances value, albeit not one that argues for a multiplicative effect per se.

As an aside: Might this be overly cautious? I certainly cannot see significance function additively. Suppose, *ceteris paribus*, that John's credence in a maximally significant proposition is .6 and Kate's is .9. Surely whatever epistemic advantage therein gained is at least in some way *proportional* to their credences. It would be very strange if the overall value they obtained were .6+x and .9+x, respectively. Holding a significant (true) belief strongly is far more likely to follow a multiplicative model instead. Now one might argue that the relationship is quadratic or exponential; perhaps one thinks that significance confers exponentially more advantage as it is believed more strongly. I don't see why, but holding this is certainly less absurd than holding that the gain is not in any way *proportional* to the credence involved.

In any case, on Whitcomb's "epistemist" picture, significance amplifies the value of belief, as does truth, justification, and whatever it is that defeats Gettier cases (from 1), as does a belief's virtuous origins (from 3), and they all do so in degrees (from 4). This view contrasts with a still very common view about epistemic value in one way already mentioned above: At its foundation, epistemism takes all knowledge-constitutive goods to be epistemically valuable *as ends*, and not as mere means—not, as it were, just in virtue of tending to bring about downstream value. Many other philosophers take it that what is of final value[95] is truth.

Whitcomb has the only reasonably worked-out account of the significance of significance on the market. Moreover, he seems to me to have a largely correct one, some small criticisms aside. Does this spell trouble for veritists? It seems to me that it does not. Nor, I hasten to add, does it pose an obvious problem for anyone else. Let's start by taking a look at the details of how Whitcomb

[95] See Korsgaard (1983) for a subtle treatment of final value.

understands the role of significance. Once we've done so I hope to be able to show how his sixth principle, the "significance amplifies value" principle can be easily extracted from his epistemism and imported into the veritist view without bringing anything contrary to veritism with it.

Here I reproduce the relative values of various doxastic attitudes compared to each other, lightly edited[96] to remove his symbolic abbreviations:

1) Believing a more significant truth is better[97] than believing a less significant truth.
2) Believing a more significant truth is better than withholding judgment about a less significant truth.
3) Believing a more significant truth is better than disbelieving a less significant truth.
4) Withholding judgment[98] about a more significant truth is worse than believing a less significant truth.
5) Withholding judgment about a more significant truth is of equal value as withholding judgment about a less significant truth.
6) Withholding judgment about a more significant truth is better than disbelieving a less significant truth.
7) Disbelieving a more significant truth is worse than believing a less significant truth.
8) Disbelieving a more significant truth is worse than withholding judgment about a less significant truth.
9) Disbelieving a more significant truth is worse than disbelieving a less significant truth.

These comparisons are highly plausible. Whitcomb justifies them by appealing to unnecessarily controversial premises, at least in one instance. 1, 2, 3, 4, and 7 are all justified by appeal to his first principle, that knowledge-constitutive goods are epistemic ends ("KCG"). 1 and 9 are justified by appeal to the "Significance amplifies" ("SA") principle. Finally, all those concerning withholding (2, 4, 5, 6, and 8) are justified by the "Withholding is neutral" ("WNE") principle.[99] I agree in the case of the significance amplifies principle. I am less certain about the withholding principle, but it seems unlikely to garner much disagreement. What of the first principle?

[96] For the original formulation, see Whitcomb (2007, 178).
[97] More precisely: "better, epistemically, as an end."
[98] As best I can tell, Whitcomb equates the value of withholding and the value of not having a view at all. See chapter 1 for a discussion of the difficulties with this.
[99] In point of fact, he seems to think the WNE principle pertains to eight of the nine cases. This is unclear to me, but belaboring the point does not seem necessary for either his or my own purposes.

If Whitcomb is right, and knowledge-constitutive goods are valuable as epistemic ends, then his conclusions follow. True belief is knowledge-constitutive in a way that false belief is not. Belief is knowledge-constitutive in a way that withholding judgment is not. But notice that we are nowhere required to appeal to the roles of epistemic virtue, or wisdom, or justification.[100] Wherever Whitcomb appeals to knowledge-constitutive goods, we can easily substitute a more broadly acceptable "true-belief-constitutive goods" principle. True belief is true-belief-constitutive in a way that false belief is not. Belief is true-belief-constitutive in a way that withholding judgment is not.

This more neutral approach should not be unpalatable to Whitcomb. Anything that is true-belief-constitutive is knowledge-constitutive, since true belief is always constitutive of knowledge (in the weak, forgiving sense of "constitutive" according to which it means "goes at least some ways towards constituting"). All this to say that we can get everything he gets in terms of a comparative schema without choosing between veritism and epistemism. No commitment to any SA view is motivated.

4.6: Conclusion:

This chapter has a negative result. Contrary to whatever we may have hoped, the small literature that exists on around the intersection of epistemic state axiology and epistemic content axiology has relatively little to show. I have argued that DePaul's cautious first premise, which holds that there are two sets of beliefs with equal numbers of members but different values, is true. Following Hu, I have argued that this is not contrary to veritism, though it may require a nuancing of the view along the line suggested. I have further argued that Goldman's bivalent conception of significant truth is unmotivated, but that nothing about veritism or any of its competitors follows.

The negative result is not entirely surprising. Contemporary philosophy has very little to say about what makes epistemic content valuable, so perhaps we should not expect a great deal about how that content relates to the much more examined question regarding epistemic state value. Accordingly,

[100] Though, admittedly, those may still function as parts of the explanation for why knowledge-constitutive goods are acquired by the agent.

starting to build a theory about content axiology may be a fruitful project. I am not in a position to know, because the project is a very daunting one, and the last two chapters of this thesis only attempt an introduction.

Chapter 5: Breaking New Ground:

5.1: Indroduction:

Earlier we saw that significance is multiplicative of value (or, more cautiously, that final value is *proportional* to significance[101]). Now we see that this principle does not commit us to epistemism. An attractive streamlining of the view would be to argue that epistemic value is at least partially a matter of obtaining significant truth, where obtaining is a matter of believing. The more significance obtained, the better. If Veritists are right, and other epistemic goods (wisdom, justification, knowledge, virtue, and so on) are valuable as means to truth, then we have an appealingly sparse theory of final epistemic value. It is a matter of obtaining truth, but where truth comes both in quantity and in quality.

I think this view is at least worth taking seriously enough to explore. In the first part of this chapter, I will attempt to demonstrate its explanatory power through a new analogy. Once I have clarified the vision, I will offer a few non-exhaustive ways of cashing it out. My hope is to demonstrate that this epistemic debate masks a deeper one that is partly a matter of psychological differences, and partly a matter of deep ontological disagreement.

5.2: Truth Puzzles:

I begin with the indulgent analogy. Epistemology might be much like a jigsaw puzzle. In my defence, epistemology is replete with such analogies, though they are usually more-or-less about either know-how or precision. One hears of maps and handbooks, and of arrows missing their targets, for instance. These are good for showing how knowledge and the like can work in many

[101] As I mentioned before, it might be the case that there is some non-multiplicative but proportional relationship between significance and value. Continua of value sometimes have inflection points or non-linear slopes. Consider the conjunction of two perfectly plausible epistemic views: 1) Credence in a true proposition must reach a certain threshold before it counts as knowledge, and 2) Knowledge is of sharply higher epistemic value than credence too weak to quite count as knowledge. If these theses are correct, then the value of my (justified, partial) belief in a (true) proposition might rise slowly as my credence rises—until I reach the point where my credence is strong enough to cross the threshold into full belief (and full knowledge), and then step up sharply. Perhaps something similar happens with significance. Perhaps the value of some bit of true belief rises until it hits some sort of threshold. That threshold might count as the boundary between totally-insignificant and somewhat-significant in a way that might appeal to Goldman (1999). There might even be such a threshold between the moderately significant and the deeply significant. Plato, on the TW view, might think so. I do not commit myself on these matters. Rather, I only want to claim that whenever the significance of a proposition is higher than another, believing that proposition is more valuable, all else equal.

circumstances. They are not particularly well-equipped, however, to deal with questions of content-based value. When the arrow hits its target, the archer has metaphorical knowledge, but we are told nothing about whether that target was worth hitting.

Here, then, is one imperfect, belaboured way of thinking about content axiology. Suppose you sit before a vast number of puzzle pieces. You want to see at least some of what a particular puzzle, the Truth Puzzle, depicts. Not all the pieces belong to it. Some are obviously not the right pieces, at least in your eyes; for other pieces, it is very difficult to say. (Some have even been constructed by partisan charlatans to have maximally similar contours to the Truth Puzzle's pieces!) Sometimes you mistakenly add these poison pieces. Some you will find, and need to rip out whole sections of the puzzle. Others will remain unnoticed.

Not all the Truth Puzzle's pieces are the same size. Some are miniscule and depict nothing you care about seeing. Some are huge and contain bits of the picture that you would very much like to see. Still others are small, but show you which of the big pieces are frauds and which are not. The big fraudulent ones you might even put aside in a second puzzle, so that you can know what those less correct than you might be looking at.

Some pieces you absentmindedly add to your puzzle (hopefully you notice them later, and hopefully you decide properly whether they really belong). Others you added after many hours of painstaking sifting. Still others you emplaced easily, after having a look at the puzzles of your betters. Many pieces fit quite tightly together (though, of course, the wrong pieces sometimes fit very tightly with the right ones). These pieces, as it were, hold themselves in place the way puzzle pieces do. A puzzle piece well-fit with its neighbours is unlikely to be dislodged.

Not all of the Truth Puzzle's pieces are good in the same way, though they are all, by definition, from the right puzzle. Some depict prosaic business, while others map out emulable depictions of flourishing human life. Sometimes overlapping portions of the puzzle satisfy your desires to see what is depicted. Some others depict great things you may or may not be particularly curious about—the origins of the universe, say, or the way time works, or the nature of love. Some

may have nothing on them that is worth the time it takes to fit them in, but may be surprisingly useful in the way they cement or dislodge other pieces that you do care about. Of course, you will never finish the puzzle. Even if you could, it might not be the sort of thing you take to be a priority. The question is how much time you should spend on it, and on which parts.

The analogy is obvious enough. In it, puzzle pieces are propositions or proposition-like entities, your puzzle is your body of beliefs, the True Puzzle pieces are true propositions, and so on. The utility of the analogy, such as it is, is two-fold.

First, we can easily see how people with different puzzles focus on different parts in a path-dependent manner. Once you've included enough of Kant, including more starts to look like it's well worth the effort—after all, Kant was a clever man, and his views on one matter, once adopted, make the rest of his views look all the more plausible. Likewise, our worldviews determine which sets are salient. If the center of a puzzle depicts something like the ceiling of the Sistine Chapel, appealing pieces that suggest atheism is true will almost certainly grab a person's attention.[102]

This is, of course, recursive. If your view of reality is disjointed and "uncentered," you're likely to believe, or be disposed to behave, in a way consistent with this uncenteredness when dealing with threats and opportunities. There is no core conviction connecting everything together. Someone whose puzzle looks more like a Mondrian painting is not likely to think there is a central "image" that bears more attention than any other part. This goes a long way, I think, towards diagnosing intuition-driven disagreements about the value of beliefs. Hardcore reductionists may feel less compelled to focus on the central image (the "middle") of their puzzles than, e.g., committed Muslims.[103]

Unrelatedly, some people think they have better things to do than their puzzles. They will want to figure out how to get on with their day, and then get on with their day. I note in passing that

---

[102] Artistic discord and intellectual discord are muddled here, for the analogy has limits.

[103] A reader points out that some reductionists are very committed to central beliefs. He suggests the example of the fanatical Darwinian who relates every bit of art, philosophy, human relationship, and so on to its adaptive evolutionary role. Such people certainly exist. However, I think a larger proportion of committed Darwinians are happy to grant that other truths are equally important than the proportion of committed theists who would allow the same regarding theism. Still, it is worth underlining that we can only speak about such things in generalities.

philosophers are probably an extremely unrepresentative sample of the population. It would be hard to find a group more interested in sorting out practically unimportant parts of life; metaphysicians and epistemologists especially.

Second, the analogy can help us think about which propositions are worth pursuing. First, there is the rather obvious practical value of certain beliefs. These are the ones that allow one to accomplish one's projects. In at least one view, this set includes those beliefs that answer questions about which one is curious. I do not agree with Hume that the value of beliefs consists in their scratching the itch of curiosity, but do admit that there is *some* kind of value in satisfying curiosity. One (at least typically) feels satisfied upon learning something one wanted to know, and feeling satisfied is a pro-tanto good thing. But let us put these matters aside for the moment.

The other obvious way puzzle pieces can help is by giving us the "big picture." But this, I think, is not easily spelled out. If certain truths are valuable just in virtue of being "deep," we cannot thereby infer that the value involved is fully cashed out in virtue of *either* merely being in "possession" of the truth of that proposition alone *or* its practical import *or* the way it is held (believing the truth gets one part of the way to knowledge). Its epistemic impact also matters. "God exists" is a proposition that, if true, surely counts as a deeply important one. But two people who learn that God exists might not gain the same degree of value. Assuming theism, a seven-year-old is plausibly capable of knowing that God exists. But compare the seven-year-old's knowledge to the case where e.g. William Rowe comes to believe theism. Rowe, prior to this imagined conversion, knows a great deal about what God *would* be like, were God to exist. His coming to believe thereby immediately grants him a great many more other true beliefs than the child is granted. By adding this one central puzzle piece he can immediately add a great deal of other pieces he knows fit with it. Over the subsequent years, still more of his view of the world will change as he reconfigures his peripheral views, but a great deal will change immediately, since he has a whole slew of new puzzle pieces, ready to go on a side table labelled "theistic hypotheticals."

I take it as uncontroversial that some beliefs, while not significant in themselves, can imply significant truths jointly with other beliefs already present in an agent's mental picture. Herein lies a

limit of the analogy. Poirot notices something insignificant about a missing piece of jewellery. As the pieces fall into place (that is, as Poirot learns the truth of more propositions), the picture emerges (Poirot works out who committed the murder). But this is still a bit of propositional knowledge, and every bit of propositional knowledge was meant to be represented by a puzzle piece. I acknowledge the weakness of the analogy in this case, but it remains a worthwhile tool for thinking about how we understand the value of epistemic content because it draws attention how our worldview shapes our inquiries, as well as the reverse.

The analogy also helps us understand the conflict between centralized and decentralized worldviews. Some people think that there is a central core to reality. Believers in the Abrahamic faiths understand God as the creator and sustainer of all existence in a robust way. We can contrast God's centrality in these religions with Olympian polytheism, which in some of its forms conceives of the gods as diverse, remote, limited—even mortal. These gods do not constitute a "core" of reality. One can learn lots of important things in a way that doesn't ultimately draw back to them. But I do not think this kind of monism-of-importance (call it "corism") is unique to theists. Mystical traditions, particularly in East Asia, conceive of some deeper reality that often does not go by the name of "God," but is nevertheless the One True Thing behind everything else (the Dao, say). And, I think, many secular people have a similar vision. I know little of Earnest Rutherford, and a quick inquiry reveals his famous quotation to be possibly apocryphal; nonetheless, there are many people whose worldviews are captured by the sentence "All science is either physics or stamp-collecting." Secular corists are frequently LaPlacian in their thinking. All that there is and ever will be can be explained with recourse to such-and-such fundamental laws—though we may not know what they are. I will have a great deal more to say on them in moment. At any rate, what unites corists is their belief that the truth-puzzle has some sort of central picture, and that if anything is worth knowing for its own sake, it is that thing.

Distributists come in various shapes as well. This is not to say there will be no regularities among them. I suspect relatively few deeply religious Christians, Jews, or Muslims think the most epistemically valuable parts of the truth are distributed very widely. Distributists believe in a

disorganized, centerless world. One corner might be very interesting—and another over there, and there—but there is no sense of a unifying picture that explains why everything is the way that it is.

I would emphasize that neither tendency strongly implies any belief or belief-like commitment to any one view of epistemic value. But it does, I think, weakly imply as much. By this I mean that corists are *more likely* than their opponents to think that epistemic value exists independently of our curiosity or even or concerns and interests, and that there is much to be gained by its pursuit. Imagine telling Plato or Aquinas that we agree with all their metaphysical doctrines about the existence and centrality of the forms or God, but that we wonder why it is that we should be interested in the Form of the Good or God. It is not that neither would have anything to say in response, but rather that their first response would likely be an incredulous stare.

Likewise, distributists are likely to be confused by strong interest concerning the origin of the universe, at least if that interest is motivated by something other than simple curiosity. Treating such questions as more urgent than questions about other subjectively interesting subjects (reptiles, French history, backgammon, and so on) might seem like an unhealthy and unmotivated obsession.

Metaphysical and ontological commitments, then, seem to be sometimes upstream of commitments regarding epistemic value. A case can also be made for the reverse, though I think it is slightly weaker. In what follows I will argue that a strong belief in corism or a belief-like commitment to corism can sometimes lead to—at minimum—the obsessive hunt for deeper meaning. It may even lead to the irrational manufacturing of meaning.

Something to notice is that, strictly speaking, the question of whether reality has a core and the question of whether all of it is connected or somehow unified with that core are distinct. There is logical space for two more views. One holds that reality has a center (where that center is the most important thing to know) but that there also exist many truths unconnected to it. The reverse view is that reality has no center in particular, but that all of epistemic reality is unified. By “unified” I mean something like “connected by implication” in the epistemic “puzzle” world and “connected by causation” in the world the real world being represented. While I have no empirical evidence on the

matter, it seems these views have relatively few adherents, in particular the corist-disunified view. Possibly some varieties of paganism conceive of outer realms that the gods did not create and do not meddle with. As for the distributist-unified view, we can conceive of those who think reality is causally joined up, but don't hold the laws joining it up in particularly high epistemic esteem. Of course people in both categories exist, but there is something awkward about the conjunction. I do not have an answer for why this is, but it seems worth noting. As the figures below suggest, the answer might be aesthetic.

Figure 1: Gustave Doré's *The Empyrean*, an integrated print with a center.

Figure 2: A section of Jean-Michel Basquiat's *Grillo*, a non-integrated painting with a center.

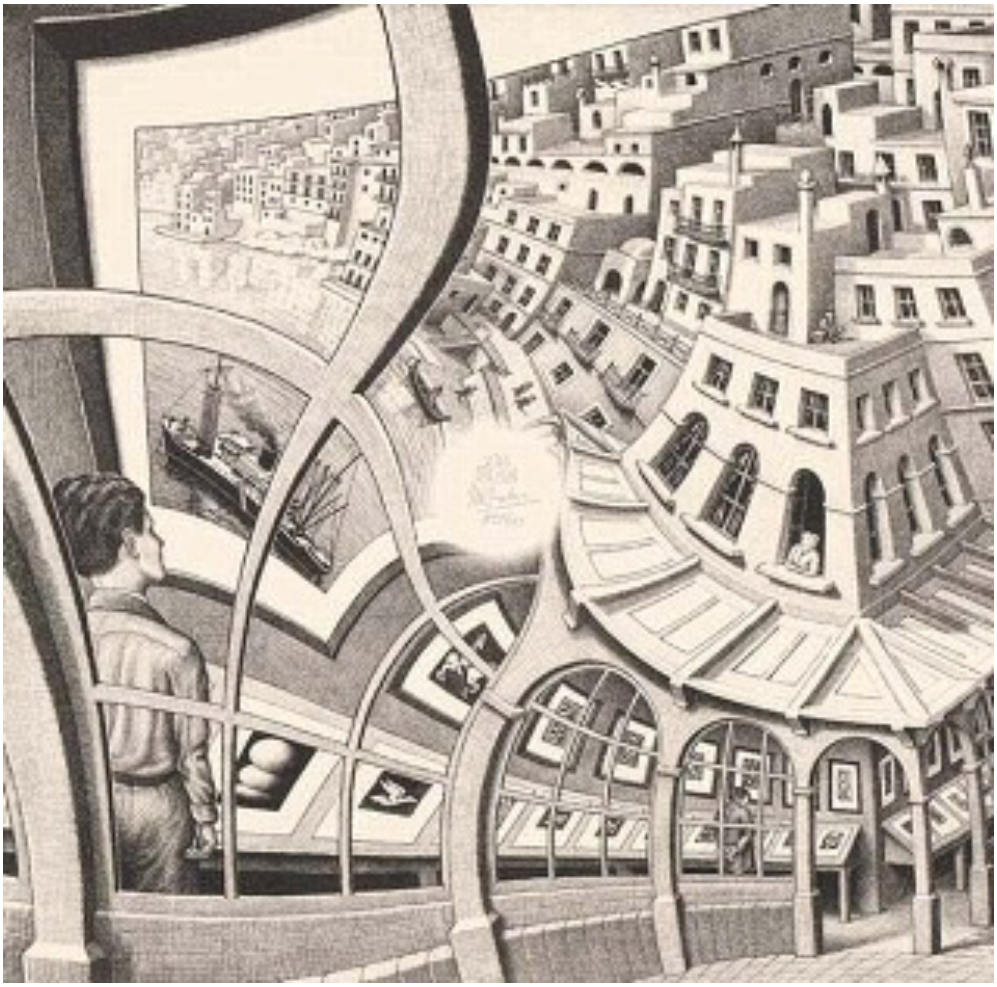

Figure 3: M. C. Escher's *Print Gallery*, an integrated print without a center.

Figure 4: Hiëronymus Francken's *Het Kabinet van de Kunstliefhebber*, a non-integrated without a center.

5.3: Hedgehogs and Foxes:

In his essay, *The Hedgehog and the Fox* (1953), Isaiah Berlin draws a distinction from Archilochus' poetry: "The fox knows many things, but the hedgehog knows one big thing" (436). Berlin's fugue on this line marks a distinction between those "who relate everything to a central vision, one system, less or more coherent or articulate, in terms which they understand think and feel—a single, universal, organising principle in terms which alone all that they are and say has significance—and, on the other side, those who pursue many ends, often unrelated and even contradictory, connected, if at all, only in some *de facto* way, for some psychological or physiological cause, related to no moral or aesthetic principle." Hedgehogs are those focussed on the center of their puzzles—imagined or real—and place pieces with the conviction that they either relate to some central pattern or figure, or are not worthy of much time or interest. Foxes, conversely, "lead lives, perform acts and entertain ideas that are centrifugal rather than centripetal; their thought is scattered or diffused, moving on many levels, seizing upon the essence of a vast variety of experiences and

objects for what they are in themselves, without, consciously or unconsciously, seeking to fit them into, or exclude them from, any one unchanging, all-embracing, sometimes self-contradictory and incomplete, at times fanatical, unitary inner vision" (436-37). I suppose it barely needs saying that hedgehogs are something fairly close to corists, and foxes to distributists. More precisely, they are the psychological analogues. Believing reality has some kind of core is, *inter alia*, a metaphysical or ontological commitment. It need not come with a personality. Likewise, having many disparate interests in the foxish way does not necessarily commit one to any metaphysical views of any sort, but it is both analogous to distributists and predispositional towards it in the weak way described above.

Berlin offers Plato, Lucretius, Dante, Pascal, Hegel, Dostoevsky, Nietzsche, Ibsen, and Proust as examples of hedgehogs; Herodotus, Aristotle, Montaigne, Erasmus, Molière, Goethe, Pushkin, Balzac, and Joyce are his foxes. This classification is not meant to be taken rigidly. Berlin admits that it is over-simple, and that if taken too seriously it collapses into absurdity. But if taken with a grain of salt, it is not merely a cocktail party game, but a useful "starting point for genuine investigation." The distinction he introduces and defends over the course of two pages. The next sixty are devoted to pinning down whether Tolstoy is a fox or a hedgehog. I have no expertise in Tolstoy and his novels are too complex to warrant arguing much with Berlin about him. I will simply let Berlin's Tolstoy be Berlin's Tolstoy.

But I will say something about (his) Tolstoy. Berlin's essay offers an excellent case study in the ways in which both psychology and metaphysics (perhaps "ideology" is better?) interact with the way we assign epistemic value. Roughly speaking, half of Berlin's thesis is that Tolstoy had a fox-like mind but an unshakable conviction that one ought to be a hedgehog: "[His] gifts and achievement are one thing, and his beliefs, and consequently his interpretation of his own achievement, another" (438). The other half of his thesis is that Tolstoy thinks we can say very little about the state of the world, and that we will never be able to say very much about the state of the world in any all-encompassing, non-trivial, way—a way that cuts to the heart of self and world—and that this is a terrible shame.

The first half is not entirely irrelevant for our purposes. It is of course true that one might have a great talent only for understanding the isolated bits of furniture, but not the room—and still

think that one *ought* to have a view of the whole room. Likewise, it is true that one might be well-suited to Great Views of the Whole, but be careless and imperceptive of the details (Dostoevsky and Descartes come to mind)—and still think one *ought* to have an elastic, disjointed vision of reality (or, perhaps, of the many bits of reality). It might be the case that being a talented hedgehog leads one to an irrational belief in hedgehogism. It certainly seems plausible that those who are frequently praised for their large, coherent visions of the whole are more likely to think it important that one have a large, coherent vision of the whole.

A further matter concerns not personality, but philosophical position. Here is how Berlin outlines the relevant part of Tolstoy's worldview: Tolstoy was deeply interested in philosophy itself, but shared his century's preoccupation with how it worked itself out in history. "History alone—the sum of empirically discoverable data—held the key to the mystery of why what happened happened as it did and not otherwise; and only history, consequently, could throw light on the fundamental ethical problems that obsessed [Tolstoy]" (444). Unfortunately, history is a failed science, beset by bias, written by fools, with too vast a subject matter, and thus one that reveals comically little. It is "nothing but a collection of fables and useless trifles, cluttered up with a mass of unnecessary figures and proper names" and it does not "reveal causes; it presents only a blank succession of unexplained events" (446). Its subject matter requires scientific approaches, but we are bad scientists, and we will never become good ones (447).

Well then, history is important, but impossible. Perhaps we ought to follow Flaubert and Woolf and focus on the interior life of the individual soul, a shift Turgenev begged Tolstoy to make, for Tolstoy could describe the inner lives of individuals brilliantly. Tolstoy, writes Berlin, "rejected this with violent indignation." For, "this was not to give the answer to the question of what there is, and why and how it comes to be and passes away, but to turn one's back upon it altogether" (451). More is impossible; yet more is necessary. We cannot be content with the small, for though we are small and the limitation of our experience inevitable, to confine one's thought to the personal is to fall into the delusion that one is in control of at least one's own small world, that one is responsible for one's successes and failures—in short that one is more than a mote blown about by implacable winds.

Yet we cannot escape our particular little realms entirely; one is condemned to think oneself free. This conjunction grips Berlin's Tolstoy very hard: "[Tolstoy] is obsessed by his thesis—the contrast between the universal and all-important but delusive experience of free will, the feeling of responsibility, the values of private life generally, on the one hand; and on the other the reality of inexorable historical determinism, not, indeed, experienced directly, but known to be true on irrefutable theoretical grounds" (458). Or, relatedly: "Freedom is real, but it is confined to trivial acts. At other times even this feeble ray of hope is extinguished: Tolstoy declares that he cannot admit even small exceptions to the universal law; causal determinism is either wholly pervasive or it is nothing… Science cannot destroy the consciousness of freedom, without which there is no morality and no art, but it can refute it" (459). Tolstoy seems to want there to be something inconsistent with his own view, something he is missing.

If Berlin is right, and I will assume he is, it is not quite that Tolstoy's mind is in conflict with itself, so much that it is in conflict with his heart in several related ways. Threads visible only to LaPlace's demon bind us individually and collectively, weaving our lives according to a pattern that very much exists, but is very much invisible to Tolstoy, and *a fortiori* to the lesser minds of the sociologists and schoolmen. Since this pattern is likely to stay invisible, we might give up looking for it and concern ourselves with smaller objects. But just as our freedom is illusory, so is the independence of these little matters. Alas, then, what will we do? The key to the puzzle lies in its great center, but that center is far beyond our meagre means.

It is important to note that this question is really (at least) two questions. There is first the psychological question, which seems to preoccupy Berlin more. When confronted with extreme epistemic difficulty, what, as a matter of descriptive fact, do we do? The second question is epistemic. If the great epistemic goods are out of reach, where do we go? These two questions correspond to the two halves of his thesis, but he is considerably more interested in the psychological question than the epistemic:

> [Tolstoy] longed for a universal explanatory principle; that is, the perception of resemblances or common origins, or single purpose, or unity in the apparent variety of the mutually exclusive bits and pieces which composed the furniture of the world. Like all very

> penetrating, very imaginative, very clear-sighted analysts who dissect or pulverise in order to reach the indestructible core, and justify their own annihilating activities (from which they cannot abstain in any case) by the belief that such a core exists, he continued to kill his rivals' rickety constructions with cold contempt, as being unworthy of intelligent men, always hoping that the desperately-sought-for 'real' unity would presently emerge from the destruction of shams and frauds… And the more obsessive the suspicion that perhaps the quest was vain, that no core and no unifying principle would ever be discovered, the more ferocious the measures to drive the thought away by increasingly merciless and ingenious executions of more and more false claimants to the title of the truth (463-4).

Berlin goes on to describe how Tolstoy longed for a theory that could survive the acid bath of his own reason, but also how he must have eventually come to suspect that he would never find such a theory. Yet, instead of turning away from the existing options, he doubles down as the angel of destruction, slaying inferior views with mounting brutality.[104]

Berlin's point is that Tolstoy knew, or ought to have known, that he would never get his hands on a true, comprehensive view of the whole—and yet pursued that end anyway. Suppose Berlin is right about this. What do we make of the hedgehogly preoccupation with the heart of things? It is unwise to categorize this tendency as a *mere* neurosis or irrational attachment, though it may be those things as well. It seems very difficult to imagine that Tolstoy would thrash so desperately towards a theory of the whole had he believed that no such thing could be obtained. Even if it were impossible, it seems that there is something very noble in at least insisting that the truth, however elusive, should not be obfuscated further by falsehoods. We think of Wittgenstein attempting his famous silence, but doing so in part by railing against the metaphysicians who would occlude things still further. One suspects that Berlin sympathizes more than a little with "foxes." Perhaps he is too dismissive of Tolstoy's destructive tendencies—after all, avoiding falsehood is a epistemically good, and helping others avoid it is socio-epistemically good.

Equally, it is hard to imagine that there is not *some* affective component driving the hedgehog back into the center of the puzzle. If we imagine that the great system-builders of philosophy were to have their work decisively refuted, their drive to theorize would remain. A Kant or a Hegel could

---

[104] Berlin notes in passing something worth repeating: There is a kind of echo of Tolstoy in French existentialism. "[They] have struck out against all explanations as such because they are mere drug to still serious questions, short-lived palliatives for wounds that are unbearable buy must be borne, above all not denied or 'explained'; for all explaining is explaining away, and that is a denial of the given—the existent—the brute facts.

never content himself with logic puzzles. Even Wittgenstein, the great enemy of Big Theory, could not manage to leave the central questions of philosophy truly behind. Perhaps he too was hedgehog of sorts, just one who, like Tolstoy, thought we could never quite get to the too-ineffable heart of reality.

Earlier I noted how some ontological or metaphysical views have implications about epistemic value. Chaotic worlds have no centers worthy of the name; orderly ones are more suggestive of a hedgehog approach. What the parable of Berlin's Tolstoy demonstrates is how plausible it is that variation in human psychologies influence human approaches to truth-gathering. Tolstoy, despite strong indications that mysteries of historical determinism will never be unravelled, could not seem to leave the question alone. I'm not convinced he was irrational to do so. If the "core" really does touch and inform everything else, as he seemed to believe, and if nothing is very intelligible without it, then it might well be reasonable—obligatory, even—to gamble on the small chance that a deeper understanding is possible. Even so, it is correct that some people are more predisposed than others to think there is a Deep Truth informing everything else, and that it is worth pursuing at great cost. This might well be neurotic and irrational.

It bears noting that most philosophers, at least historically, have been hedgehogs. One reason for this is that philosophy is a costly activity, and people are unlikely to pursue it when other less costly epistemic endeavors (or non-epistemic endeavors!) are cheaper, unless one has *some sort* of conviction (I use the word loosely) that there could be a worthy payoff. In the time it takes a person to wrap her head around Kant's metaphysics she could learn much more about any number of interesting things.

Another reason concerns consistency. A fox, as I understand the term, is far less likely to bothered by her various beliefs contradicting each other. This is not to imply that those not preoccupied with finding a center are irrational or stupid. Rather, they are simply are simply less likely to notice. Suppose an evangelical Christian of a foxish mindset believes that the Christian scriptures are more-or-less literally true. I have no problem believing such a person could not also be a very clever geologist, for example. If the conflict between the literal claims of the Pentateuch and the carbon dating of rock formations is pointed out to her, we can imagine a sort of shrugging off—for

it is easy not to have one's worldview shaken to the core when one doesn't believe in such cores. She can go back to working on her favored bits of puzzle without feeling too threatened. So what if there is an inconsistency? Perhaps some correction here or there is needed, but there is no great need to reorient every bit of the puzzle. For, in the heart of the fox, the epistemic payoffs are distributed across great distances, not amassed in one place. If Quine is right, and our beliefs form webs, it is not to everyone intolerable that there exist tears between the web's center and its periphery.

The hedgehoggery of the philosopher may have worrying implications for philosophy. A Plato may have some not-entirely-rational need to believe in the centrality of the forms; an Aquinas may need the most valuable kind of knowledge to lie in the heart of a reality that may turn out to be heartless. Even if the hedgehog turns out to be right about the *existence* of a core, she may be tempted as Berlin's Tolstoy was tempted—to waste her life looking for what she can never find. As I have mentioned in passing, my sense is that those philosophers who are aware of the danger nonetheless find themselves impelled to behave a lot like hedgehogs in their anti-hedgehoggery. For a man who thought that metaphysics was largely nonsense, Wittgenstein certainly spent a lot of time thinking about metaphysics. I confess an intense conviction that, whatever turns out to be real is the sort of thing that hangs together marvellously around a small number of central truths, though I worry the conviction is not proportioned to the evidence.

If objectivity is important, and it is important, then the philosopher must be careful not only to understand the implications her other commitments have for the structure of epistemic value, but to be on the lookout for how intensely she feels the need for there to be great stores of epistemic value at the end of the philosophical quest to systematize reality. Those who mock philosophers tend to portray them as fools and obsessives set on finding some obscure and dubious truth. The fact that these who mock are not usually philosophers is insufficient for their dismissal. A particular type of person does philosophy, and that person seems to be the same sort of person who believes in a core, regardless of whether she thinks she's very likely to find it. The French existentialist referred to above, who believes existence must be suffered without satisfying answers, risks missing her own

unmotivated commitment to The Question so dignified that it demands we suffer for it. That dignity she does not get for free.

As I said, I have an unshakeable hedgehogly conviction that there are deep questions, and that answering them brings a distinctive value independent of my curiosity or my interests. My reader may or may not share this conviction, but if she does, I hope I have gone some way to showing why someone might think the way Hume does. His position seems less absurd from the vantage point of the distributist; distributism might seem quite plausible to foxes, and many people are foxes. Of course, I do not want to imply that this is the only causal sequence in operation. My hunch is that people's fox or hedgehog tendencies are more commonly the product of their beliefs about the world than the reverse. Jointly or separately, in one order or another, they produce convictions about whether propositions have content-based epistemic value independently or merely because of our desires and interests.

I do not claim to have teased out all distinctions pertaining to grappling with CA. Rather, I have offered partial explanations for why the question is so difficult—indeed, for why it is so difficult to know where to start. In the next section I raise more. However, I do not think these sorts of considerations are fruitless. Ground-clearing is of course necessary for philosophical progress, and drawing battle lines useful. I hope to have done slightly more than that, however. There are tools available for getting past the incredulous stares Humeans and Thomists give each other. Once we realize how entangled metaphysics and ontology are with questions about content axiology we are less likely to spend all our time debating the status of various epistemic states, and are (possibly) more likely to make some progress.

5.4: Yet Further Difficulties:

In this section I will address two methodological difficulties. My goal in this section is to get a sense for what is epistemically valuable in itself, if such a thing exists, rather than what might be merely practically useful or enjoyable to think about. In the first part I describe what I take to be the intuitive way to get a sense for that which has objective epistemic value (at least to contemporary

analytics). In brief, the idea is to create a list of truths that seem broadly likely to be deep and then extract general principles inductively. Nevertheless, there are more perils that need to be pointed out. They are not particularly surprising ones, given the scope of disagreement historically. The second difficulty concerns teasing apart the "objectively" or "non-derivatively" epistemically valuable from the practically or instrumentally valuable.[105]

Kripke (1980) offers a useful method for the early stages of conceptual analysis. If we are interested in understanding X, we should ensure that our ways of talking or thinking about X actually refer to X. It is not necessary that we have a perfectly functional definition of "X," our term for the thing in the world, in order to refer to X. We just need something good enough to fix the reference. That can be a single property; even one that, in the final analysis, ends up being an inessential property of X. In his illustration, we can fix the reference of our term "heat" to *heat* by pointing to that which causes in us a certain kind of sensation. It may turn out that in a different possible world heat does not cause such sensations in us (more properly, in our counterparts), but this would only be a problem if we were already seeking, at this early stage, to understand heat's essential properties or undertaking some similarly definitional endeavor. Rather, we use the property of causing-such-and-such-a-sensation to fix the reference of our term; after that, we can keep or dispose with the property as needed. Either way, right or wrong, we'll still be talking about that thing that feels hot.

This, in any case, is how Sorensen (2011, 662-3) thinks we ought to begin when trying to determine what is interesting. Just as heat is an objective property out there in the world, so is what is interesting. And, just as we use the feeling of heat to come to grips with the world, to ensure that we're at least talking about the same phenomenon out there, so we can use collections of what smells interesting to get at that as-yet-uncomprehended objective interestingness. I think Sorensen is right about the objectivity of the interesting. It is indeed intuitively persuasive to observe that some numbers would be more interesting than others even if human beings never existed (663). But how the

[105] There is some work already done making this distinction more precise. It is not necessary for our present purposes. In this section, I am not so much interested in defining exactly what objective epistemic value is, but in gathering a set of examples of objectively valuable truths that is as untainted by practical or derivative value as possible. Once we have wrestled with some cases, I wager, much of the important clarificatory work will already be done.

interesting relates to the deep or important is a difficult question. At least on a superficial level, some interesting-looking facts seem pretty unimportant, though it is less apparent that what is deeply important could ever be objectively uninteresting.

There is nothing in principle wrong with the Kripke-Hare approach. Nonetheless, it seems that building a widely-acceptable list of interesting examples from which to extract principles is far more difficult than anyone has heretofore said, mostly because judgments of depth in particular (and interestingness to a lesser degree) are unavoidably slaves to our broader commitments. I don't suppose there is anything particularly odd about this. It is widely recognized that, loosely speaking, everything in philosophy has implications for everything else. In my estimation, this is why there is so little work done in analytic philosophy about content-derived epistemic value. Philosophers sense that there is no single broadly-accepted system for pointing out where the deep truths lie; rather there are many such systems, and subdivisions within those systems.

I contend, then, that what is epistemically important depends on ontology and metaphysics. What ends up existing (and how it exists) has broad implications for the epistemic treasure-hunter. It is worthwhile to discuss the example of human action. How "deep" are the truths about human beings? Aquinas thinks humans are very important. On his view, there are no other intelligent animals in the universe, the universe was created largely for humans, and the functioning of the universe was tailored to human ends. A more common view in contemporary philosophy is that there may be many forms of intelligent life, that the universe was not created at all, and that the functioning of the universe is indifferent to human *teloi*, insofar as *telos* exists at all. An extreme skeptic, to take a third case, might think that she is the only intelligent being—some ephemeral Boltzman-brain momentarily emerging out of the endless roiling chaos. All three philosophers will agree about their own existence, and maybe even about some of their own important intrinsic properties; none will agree about the relative importance of facts about themselves.

What may undergird this disagreement about the importance of human truths is disagreement about the human place in the structure of reality. Compare the imagined contemporary philosopher's disagreements with both. It seems to me that she is likely to have views closer to the radical sceptic

than to Aquinas, at least if she accepts that there is objective value to epistemic content at all. This is not because she denies that human beings have moral worth, or because she values human relationships less; on these matters she is likely to be far closer to the Thomist than the radical sceptic. Rather, she is likely to think that human beings have less to do with the architecture of reality than the Thomist does. For her, the heavens are equally unperturbed by our presence—and it is their movements that cause everything to be how everything is, more-or-less anyway, rather than ours.

No serious person thinks that human beings compose more than a miniscule fraction of the mass of the universe, but many serious people think that knowledge of human workings is a very important kind of knowledge, even, I would wager, "objectively" speaking. One view is that this is because the latter group places human beings towards the center of the causal structure of the cosmos. I stress that I do mean *causal*. One might very well think that human beings are pretty much all that "matter," that everything else is just so much dust and rock, but it is only if one thinks that the cosmos was in some way designed around human beings that one has license to think the truths about humans are deeper than the truths about the origins of humans. Certainly there are several senses of "cause" at work; no one thinks human beings created themselves or the universe. But some reasonable people, not least many theists, do think that the world was designed for humans, or something of that sort.

Will many naturalists agree about the objective value of knowledge of human beings? My guess, and it is only a guess, is that those naturalists who do agree are likely to agree on the basis of human distinctiveness. Some naturalists believe that human beings are the only intelligent beings in the cosmos, and furthermore that many deep truths are about intelligent beings. Naturalists who are also property dualists might think that quite a high proportion of the kinds of things that exist only on Earth. Without human beings (or conscious beings in the case of the dualist) there would be much less reality, as it were. Naturalists with much lower views of consciousness or who think humans not particularly unique are, on this view, much less likely to privilege knowledge of human affairs in this way. This, I suspect, is broadly what we would find empirically.

What then do we say about the existentialist archetype? I do not intend to delve into Sartre or Camus. I merely mean to note that there are plenty of philosophers who think that the universe is a

cruel joke that vomited up human beings without meaning to, and, for all that, that human beings are the worthiest subjects of study, and human truths the most important to access. It would be patronizing to insist that these thinkers seek to know human beings only because knowledge of them is derivatively important for achieving our ends (though, of course, it is that too). At least, this does not seem like the fairest way to characterize this line of thought. A better way might be to see this stance as rejecting the need for a certain kind of objectivity in the first place. Only with human minds can there be meaning and valuing, even if the meaning and valuing are somewhat arbitrary in content. If the universe is fickle to the point of cruelty, and random to the point of being inscrutable to human eyes (eyes laden with concepts to which the physical universe is unintelligible), then perhaps turning to the human world as the only one we can properly grapple with is all that we have left.

Contemporary analytic philosophers are less likely to see things this way. For many, though the physical universe was in no important way built around us, we can still meaningfully come to grips with the bedrock truths by exploring the natural sciences. Many on both sides of the Channel might agree about all the ontology and metaphysics of the natural world and our insignificant place in it. But this may not yield agreement about what counts as deep. I draw attention to this to underscore the limitations of our empiricist approach. We are not guaranteed to agree about what is deep or interesting or objectively worth knowing *even if* we more-or-less agree about the general metaphysical architecture of the world. Non-analytic philosophers are far more likely to think that the world of human thought and emotion cannot be exhausted or reduced to those objective inquiries of the natural scientist or the reductive metaphysician.

Yet, for that reason, it may be, after all, that there is a fundamental disagreement about what counts as real in the first place. The existentialist admits that his analysis of self does not carve out a metaphysical space for the self that is comparable in its hardness (or facticity, or verifiability) with the subject matter of chemistry. Heidegger knows that a table and an altar may be physically identical (made of stone, of such-and-such dimensions, etc.), but he denies that they are identical on even the

most basic level. A continental thinker may be no substance dualist, but she still might think that there is a level of human reality unanalyzable in analytic or scientific terms.[106]

Of course, all her views on this matter may be wrong! Perhaps she is confused about what she is doing, and has mistaken psychology for philosophy. What I am trying to point out is how tangled the epistemic and metaphysical questions here are. I began by trying to identify obviously deep matters, in order that we might abstract, through some kind of induction, what the features of epistemic depth are. Then we saw how judgments of epistemic depth are dependent on metaphysical commitments. But it seems that these, in turn, require epistemic disentanglement. Do the existentialist and the garden-variety analytic reductivist disagree about the architecture of the world in an important way? It is not immediately clear, and would not become so without much, much more enquiry.

As I have said, there is almost no recent work on this question in contemporary analytic thought. But what work there is has picked out a favorite example of what must count as uncontroversially important, objectively valuable subject matter: scientific laws. Below are a few examples from the last two decades.

> *It is better, epistemically, to believe* significant *and deep truths about e.g. science or metaphysics, than it is to believe* trivial *truths*. Whitcombe (2007: iii).
>
> *Any account of the interesting should at least try to respect our judgements about the [cases of] fundamental scientific laws [and] the existence of God*. Grimm (2011: 521).
>
> *Suppose… one is faced with a straight choice between two propositions… one of which is profound—a fundamental scientific truth, say*. Pritchard (2014: 122).
>
> *Consider the following two beliefs: 1. Someone's belief that the universe is expanding at an accelerating rate. 2. Someone's belief that the number of people ever to have visited the David Hume memorial up until now is even.* [In what follows, he supposes that 1 is obviously more valuable.] Hu (2016: 9).
>
> *That is, even from a purely epistemic point of view (i.e., setting aside the practical utility of being aware of the deep scientific truth), it seems that one should value the significant truth over the trivial one.* Pritchard (2021: 5517).

God and metaphysics get mentioned, but the least controversial significant truths have to do with the laws of the natural science, presumably with physics as their queen. Now, in fairness, it is plausible

[106] This is not even to mention figures like McDowell or Charles Taylor, who do not fit my stereotype of the analytic in this context.

that most thinkers today who acknowledge objective epistemic value would think the laws of physics are reasonably significant. But not so for Plato or Aquinas. Naturally, some part of the explanation for the difference lies with the fact that roughly three quarters of contemporary analytic philosophers are atheists;[107] presumably, Plato's Forms are even more unpopular than God is. But what of the prominence of physics? In my own eyes, learning that the universe has been on average expanding throughout its history seems a truth of lesser significance than many.[108] Is it obvious to everyone else that this is a paradigm deep truth?

I can think of at least one reason to doubt it. The truths contained in the works of St. Paul or Burke or Tolstoy are plainly the sort worth knowing, but they are deeply bound up with the world of human action. On a sufficiently austere picture, these might count as merely subjectively significant. Deep in their way, perhaps, but only insofar as they help us get through our lives. Somewhat stranger is the fact that (with one exception noted above) philosophers do not seem to be very confident about the objective significance of philosophy. Even this can perhaps be explained as a concession to reductive materialists. After all, it is still fairly popular to think that all truths (in some way, in the end) reduce to physical truths.

If this line of argument is right, we ought not read the focus on the scientific as an emergent consensus that the truths of physics are the deepest; rather, that the truths of physics lie in the intersection of sets "widely accepted to exist" and "least likely to boil down to something merely subjectively important." I think the (admittedly non-unanimous) omission of God provides further evidence for this. *If* the Abrahamic God exists, and *if* anything is worth knowing about in its own right, it's certainly that God. Even the most strident atheist is likely to agree with the conditionals. Likewise, if there is no God, it isn't the case that knowledge of God is significant because knowledge of God is impossible.

---

[107] https://philpapers.org/surveys/results.pl

[108] At least in itself. Sometimes this fact is used in arguments for God's existence or against the existence of a past eternity. These seem more important.

This offers a useful methodological tweak to the Kripke-Hare method for further pursuit. Instead of looking for what we find significant, we ought to look for conditional significance, which is more or less what we have been doing already: the Forms for Plato, God for Aquinas, and the laws of physics for the reductive materialist. Unfortunately, this is not a hugely surprising conclusion. It is, I admit, exactly what one would have guessed. Still, we can perhaps be all the more confident in this picture for this reason.

One important upshot might not be so obvious, however. It might be tempting initially to think that the deepest question, or at least one of the deepest questions, concerns what exists. The above considerations certainly provide plenty of motivation for trying to answer this question when looking for the deepest truths. As I have shown, it is impossible to get very far with a "neutral" non-committal approach to metaphysics. The options are too numerous, and the consequences of choosing any one are far too weighty. But this seems to be primarily a methodological difficulty. We must find out what is real before we can try to get to the heart of reality—at least on the views we have looked at so far.

But one might reply: There is a distinction between "the deepest truths" and "the set of significant truths." No serious person would ever think that e.g. having true belief about this or that phone number is as objectively significant as having true belief about the fundamental causes that shape physics. Of course, this line is partly correct. Absences are not in themselves significant, so whatever is significant must be included in the answer to the question "What is there?". Nor do I think it is true (or at least obviously true) that knowing about mundane things is totally without objective value. Rather, it seems that getting a laundry list of the things that exist does not seem to have satisfied philosophers in the past, though it's a good place to start.

5.5: Conclusion:

The reader might agree with my intuition that knowing some parts of reality is more valuable than knowing others, independently of facts about us. If so, I think it likely that her cosmos has some kind of center. Perhaps I am mistaken about this; what I am confident of, however, is the feedback

loop between our ontologies and epistemic axiologies. This may explain why contemporary philosophy says almost nothing about CA. Because our discipline is diverse and fragmented, there is little agreement about basic philosophical questions. It thus becomes extremely difficult to propose a view to a general audience without eliciting many incredulous stares.

What I propose, then, is an epistemic fragmentation appropriate to a diverse discipline. We should never have expected people with stridently divergent metaphysical views to agree about what parts of the cosmos are worth learning about. Going forward, we ought to tolerate e.g. work from moral realists about the significance of moral truths that anti-realists find dull and implausible, and vice versa. Yet, for all that, I suspect intuitions about depth will be surprisingly gripping for some. The Hedgehogs among us may have anti-corist views that sit poorly with their deep sense of what is worth knowing; the same goes for foxes. Dragging these intuitions to the forefront may, I hope, provoke internal confrontations that yield more consistent “big pictures” in more than a few puzzles.

Chapter 6: Strengthening Subjectivism:

6.1: Introduction:

My aim in this chapter is to give a defensible subjective approach to epistemic value. "Subjective approach" here means one that indexes value wholly or largely to us, rather than the world. That is, this view denies that some content is valuable in its own right, independently of how useful or interesting it is to us. While I think this view fails, I argue that it is reasonably good at avoiding some strong objections. Precisely labelling the view in question is difficult, for it is a kind of hybrid combining moderate, pro-tanto Humeanism with reductive pragmatism. My strategy will be to nuance each, then to combine them into a two-level view that preserves the virtues of each while curbing their worst excesses. Here are some low-resolution definitions from which to begin:

> Humeanism: Epistemic value is a distinctive kind of value obtained when curiosity is satisfied.
>
> Reductive Pragmatism: Epistemic value is neither final value, nor intrinsic value. Something is epistemically valuable just in case it is an instrumentally valuable means to a practically valuable end.

These views are problematic, for some reasons we have already seen (and some I will say more about). Humeanism has the following problems:

1) It allows for too much epistemic neuroticism. Sometimes we are curious about worthless questions (e.g. about the name of the song playing in the elevator yesterday), or out of false beliefs or ignorance.

2) It does not easily make sense of cases of boring questions (that do not provoke curiosity) having interesting answers, or of interesting questions (that do provoke curiosity) having dull answers.

3) It is on dubious normative grounds. Since the fundamental good being targeted is curiosity-satisfaction, why be so concerned with truth? Moreover, curiosity-satisfaction

seems like a relatively cheap good to begin with. It is hard to imagine strong independent norms developing here.

Reductive pragmatism has not-too-distant problems, though, since it is so unpopular[109], I will not examine them in detail:

1) It also gives very counterintuitive results, offering no obvious way to sanction most intellectual passions or hobbies; great revisions to our common-sense approaches to these will be required. It is not clear pragmatism even sanctions many or most so-called "deep" questions, which Humeanism can at least recognize as typically interesting.

2) It threatens to blur out of existence the well-recognized distinction between the epistemic and the pragmatic. Doing so is highly revisionist, and perhaps threatening to the boundaries of the discipline—where those boundaries are "load-bearing," part of what it is about epistemology that tracks something in our language and in the world.

This list is not exhaustive, but it gets at some of the core problems for subjectivists.

After I have examined the problems with both species of subjectivism, I will argue that the most defensible position is a moderate, nuanced kind of pragmatism, but one that includes a healthy respect for Humeanism. To give the reader an inkling where I'm going, here is a brief articulation of the view:

Moderate, Two-Tiered Subjectivism:

The fundamental grounds for epistemic value are pragmatic, insofar as the reason we should care about epistemic value is that a healthy epistemic life provides a partial basis for a life well lived. However, on a lower level, the best way to have a healthy epistemic life involves deference to certain desires (or desire-like phenomena) like curiosity. Even if an individual tokening of curiosity-indulgence is itself neutral or even counterproductive with respect to our practical welfare, given human limitations, we ought in many such cases to defer to our

[109] Wrenn (2017) is an exception. He defends the view that truth has no (or almost no) intrinsic value.

> curiosity, for several reasons. Doing so helps sustain a flourishing intellectual life where intellectual asceticism might, in the long run, lead to stupor or exhaustion. Further, learning about what is "interesting" will, for more-or-less evolutionary reasons, lead to practically helpful knowledge.

In this view, pragmatism provides the bedrock justification for caring about epistemic value. Epistemic value itself is all-things-considered valuable insofar as it is instrumental in helping us live well. However, given how motivating curiosity is, the goals *within* the *epistemic value system* are determined in large part by curiosity. It sustains interest for individuals and helps diversify and thereby strengthen our communal stock of knowledge.

Two immediate caveats: First, the careful reader may worry about the consequentialist presumptions this talk of "values" imports. This is a fair worry. Many epistemologists think there are epistemic norms that diverge from what is implied here; that is, they think there is more to epistemology than garnering epistemic value. For instance, one might think that law-like norms about avoiding all possible error trump the acquisition of valuable knowledge. Relatedly, strong reductivists may think there are *no internal* positive epistemic norms about garnering truth or knowledge. These debates are worth pursuing, but I will not pursue them here. This thesis's primary aim is simply to get a handle on epistemic value. As for why positive norms about satisfying curiosity must be *internal* to the epistemic level, my hope is that the answer will emerge by the end of the chapter. In brief, the idea is that our epistemic lives are shrouded by our psychological ignorance and suffused throughout with complexly interwoven, frequently intractable yearnings, and any attempt to tinker with them too much risks making us worse off.

Second, given this thesis's focus on epistemic value, one might think it unusual to introduce the question of the moral (or all-things-considered) justification for epistemic norms—especially this late in the game. But the moral dimension of epistemic value is the major advantage subjectivism has over objectivism, and philosophical charity obliges me to include this advantage. Objectivists can of course tell a story about why epistemic goods matter, but there is something offputtingly mysterian about appeals to value "for its own sake." Is the intrinsically valuable kind of knowledge also morally

valuable? Is it valuable all things considered? If not, in what does this non-derivative value consist, and how ought it be weighed against other kinds of value? The answer is not obvious. In contrast, subjectivists have a clear story available about why we should care about epistemic value at all doing so contributes to an uncontroversial kind of value, the value of living well. I have little to say about this difficult question, but simply wanted to mark the subjectivist's relative advantage in answering it.

6.2: Problems with naïve Humeanism:

I now return to the task at hand. In this section, I will lay out some worries about Humeanism that motivate revision. Now, I take it as uncontroversial that there is some connection, possibly a very loose one, between the value of some truth and our being interested in it.[110] It would be a bold view indeed that denied *any* connection between what we are curious about and what is valuable. But what does the connection look like?

In both the nuanced subjectivist view I describe in this chapter, and my own objectivist view, this connection can be preserved without curiosity itself bestowing epistemic value. Rather, it might be that we are (frequently, or at least sometimes) curious about something because our faculties track what is valuable. This could be in virtue of simply being epistemic agents with minds oriented (more-or-less well) to tracking down significant truth either for its own sake, or because it is constitutive of human flourishing, or because doing so aids in our other endeavours. In naïve Humeanism, by contrast, the causal connection goes the other way. It is not that our true beliefs are valuable because we are good at tracking down valuable truths, but rather our true beliefs are valuable because our being curious about their subject matter renders them valuable. Sosa (2003) holds:

> To want the answer to a question, for its practical value or simply to satisfy our curiosity, is to want to know the truth. If I want to know whether *p*, for example, I want this: *to know that p, if p*, and *to know that not-p, if not-p*. And to want *to know that p, if p*, is to want *to know that it is true that p, if it is true that p*; and similarly for *not-p*. So our desire for truths is largely coordinate with our desire for answers to our various questions (157).

[110] See Brady (2009, 266) for a list of philosophers who agree that the connection is reasonably close.

As Sosa says, this view, like every other view, is compatible with our believings and knowings *also* being of practical value. Humeans are looking for an additional kind of value, one bestowed by curiosity satisfaction.

Grimm's "What is Interesting?" (2011) will form the backbone of much of the view I here elaborate. In a previous paper, he implicitly argues for a kind of two-level view like the one I advance. Prior work even emphasizes the shortcomings of Humeanism. He diagnoses one Humean we have already seen, Goldman (1999), and another, Hempel (1965), in the following way:

> According to this way of thinking, our curiosity about how things stand in the world is therefore importantly like the thirst we (characteristically, at least) feel when our body is dehydrated. When our body is dehydrated—when we experience thirst—*satisfying* our thirst is naturally thought to possess a kind of intrinsic value: it seems to be a good in its own right, quite apart from whatever further contributions it might make to our well-being. The fact that we are thirsty, moreover, seems to provide us with a standing or *pro tanto* reason to satisfy our thirst, a reason that holds in virtue of the intrinsic value that comes from satisfying our thirst. Of course, that is not to say that satisfying our thirst does not importantly contribute to our broader well-being; in satisfying our thirst (poisons aside) we replenish our body in a way that helps us to perform our everyday tasks. The point is simply that satisfying our thirst seems to have value in its own right, over and above these other contributions (2008: 727).

This is a familiar Wittgensteinian move, and an appealing one.[111] There is a clear difference between desire satisfaction and desire removal; the former bears an obvious appeal over the latter. That preferences can be satisfied with less than "real" fulfilment is not, thus, the best line of attack against the Humean. The better line concerns what happens when curiosity goes awry. Brady (2009) offers some cases that help illustrate this:

1) Curiosity downstream of false beliefs:

   Suppose some charlatan has convinced you of the power of crystal healing. Given how interesting crystal healing *would be* if it *were* medically sound, you might become very curious about what the experts on crystal healing say on specific matters. You might easily satisfy your curiosity with truth—for one can easily come to true beliefs about what they think even while in error about the truth value of the content of those beliefs (270).

[111] *Philosophical Investigations, §22*. Sosa (2003, 157) argues along the same lines.

2) Neurotic curiosity:

Sometimes we find ourselves wondering about questions we ourselves know to be without value. Brady offers the example of step-counting or wondering about the ratio of blue to red books in the library. Perhaps a more common example would be wondering about some detail from the distant past, like the name of a restaurant one went to on childhood vacations or some bit of gossip from the distant past. For Brady, "there seems to be something objectionable about the idea that the truth has value because it answers the subject's inquiries or questions, in those instances where the subject does not *value* getting to the truth of those questions" (271).

These are not the only worries for the Humean, but they are worrying enough to demand answers.

Here are two responses available to the Humean:

1) Idealization:

Errant desires are common in all areas of philosophical psychology, but they don't typically show the worthlessness of the relevant *kind* of desire. An 18th century French nobleman named Tarrare desired to eat everything, but this does not show that satisfying hunger is without value, for we can imagine what Tarrare *would have* desired had he not suffered from mental illness. And so with curiosity. We can posit that curiosity satisfaction is valuable if and only if a sufficiently rational and informed agent would desire to know the truth at hand.

2) Omni-curiosity:

Some philosophers (notably Lynch (2003) and Kvanvig (2004)) hold that we have a standing, open-ended interest in knowing the truth. Neurotic or misguided curiosity, on this picture, involves a kind of normative failing in that it wastes time and energy better applied to other pursuits, either practical or intellectual. Nonetheless, because we have an open-ended interest in knowing the truth, our errant curiosity still lays hands on

> something of some value insofar as it gets at the truth—otherwise problematic though that curiosity may be. As Brady puts it: "If we assume, then, that we are naturally curious about the truth *simpliciter*, then the worries about how to close the gap between our valuing the truth disappear" (272).

Regarding the idealization strategy, it is difficult to imagine the sort of ideal agent the Humean needs. Idealized curiosity is curiosity free from the wrong sources, namely curiosity-provoking false belief and neurosis—but, as Brady points out, she cannot be "provided with too much information," or else "it is more than likely that her curiosity will disappear" (272). So she would have a combination of true beliefs and ignorance about what we can call, somewhat roughly, interesting subjects, and she would be untroubled by niggling neurotic musings.

In many cases, however, we can be led astray not just by false belief, but by ignorance itself. Assume that atheism is true, and imagine an agent who thinks atheism and e.g. Judaism are equally likely, and that no other religion or religious stance is likely at all. Now, atheism is interesting and (if true) of obvious epistemic importance. However, because it implies less than Judaism,[112] it would be less epistemically important, if true, than the truth of Judaism. She will be very curious about Judaism and its implications for the rest of her worldview; probably, she will be more curious about e.g. what its scriptures and traditions imply than she will be about what atheism implies. This is because Judaism (again, if true) implies a great deal more. It *already* implies a great deal for her in a conditional way. If rational, she will think the constitutive teachings of Judaism to have a likelihood of roughly half. If atheism is true, as we have assumed, working out Judaism's implications is not the best way for her to spend her epistemic resources. So false beliefs are not even required for the kind of mistakes we would want an ideal agent to avoid.

Some sceptics might be unconvinced by the inherent "interestingness" of religion. But it is easy to think of other cases. Most people are agnostic about the existence of intelligent, sophisticated

[112] Judaism, like the other great world religions, has a great deal to say about ethics, for example. Atheism (by itself) is relatively quiet on the matter. Of course, there may be interpretations of Judaism that diverge from my analysis; if so, the reader can supply the ethically prescriptive religion of her choice.

life outside of Earth. But compare how most people would react to the news that some scientist proved that such a race of aliens exists with how they would react if someone proved that they do not. I grant that both scenarios are interesting, but *surely* the former case would provoke far more curiosity. Now, again, supposing there are no aliens, it is likely that agnosticism about aliens provokes the wrong kind of curiosity in the person who wonders what aliens would be like, given the (50% subjective chance, say) that they exist.

Yet, suppose all these difficulties can be surmounted–after all, the mere fact that we don't know how to fill in the details of a highly rationalized, idealized account of curiosity does not make that view false. It seems to me that such a highly sophisticated, but practically inaccessible account risks losing a great deal of its plausibility *because* of its sophistication. On such a view, we could never say with even remote confidence that a given instance of curiosity-satisfaction gets at any value. There is something very worrisome about combining this claim with the claim that curiosity is what makes some true belief epistemically valuable. For any token inquiry, the best one could do is to say that a semi-divine (but still somewhat ignorant!) sage might be curious about that subject, and that it thus might be the case that one is engaged in something epistemically valuable. But, I take it, part of the bedrock appeal of the Humean view is that it appeals to the satisfaction we get when we get what *we* want—not when we get what some utterly remote imaginary sage wants. My own (rather tentative) view is that the epistemic sage is just as good at asking the right questions as the moral saint is at doing the right things—that is, very good and very worth emulating—but that the varying value distributed across various truths was, as it were, already there. If one thinks that curiosity itself bestows value, then the vast distance between the sage's curiosity and our own renders Humeanism esoteric, and intuitively unappetizing.

I turn now from idealization to another strategy to defend Humeanism, the argument that we have a standing omnivorous interest in truth as such. Kvanvig claims that interests are typically (perhaps always) general:

> [We] do not have an individual interest in the truth of the claim that our mothers love us, that the president is not a crook, that Wyoming is north of Mexico, and so on. What we have is a general interest in the truth, and that interest attaches to particular truths in the manner of

> instantiation in predicate logic. The default position for any truth is that our general interest in the truth applies to it (2003, 41).

The first thing to observe is that this stretches the definition of Humeanism to the breaking point. Plausibly, we have an interest in knowing every truth. But we are certainly not curious about every truth. One could argue that we have some kind of general positive disposition towards the truth, of course, but that no more constitutes curiosity than my general pro-tanto interest in not getting hurt constitutes fear. Fear is not an outlier among affective states, as Brady notes:

> [What] is true of fear is true of other emotions, such as disgust, shame, anger, guilt, joy, surprise—and curiosity. Each of these emotions has the role of alerting us to something significant or important, and fulfils this role by directing our attention onto certain objects and events and not others. Curiosity, like other emotions, involves selective attention (274).

Second, it is not clear that we do have a standing, open-ended interest in the truth—or, at least, that is not the best way to put that claim. Earlier I laid the foundations for a hierarchical objectivist view, according to which some parts of reality are perhaps more central and more important to know than others. If that is right, then it might be the case that truth is the kind of thing in which we are interested, and that if we had infinite capacity and time, we would do (epistemically) well to acquire as much as possible. Of course, since we are limited in all sorts of ways, it makes sense that we would prioritize. From an economic standpoint, it makes no sense to spend an hour fishing a penny out of a fountain. Life is too short for such a small reward. A hierarchical view in epistemic axiology is one way of making sense of what Kvanvig gets right—that we have some kind of standing interest in knowing what the world is like, without committing to anything radical.

Above I have argued that Humeanism has trouble with the wrong kinds of curiosity, namely curiosity stemming from neurosis, false belief, or some kinds of ignorance. To my mind, this is the biggest problem with the view, but there are others worth mentioning. Consider cases where we are not curious, have no reason to be curious, but miss out on important or interesting truths, and cases where we are curious (and have every reason to be) but receive paltry rewards. There is little need to list examples; two will suffice.

Gold under dirt:[113]

Marie finds herself working on the history of potato prices in her very dull job for an investment firm. She discovers that potatoes were once worth their weight in gold during the Alaskan gold rush.

Dirt under gold:

Claudio, a Renaissance alchemist, discovers grains of gold in his crucible after an intricate investigation based on an enthralling manuscript. He investigates further and discovers that a local miscreant had placed the gold in there the night before as a prank.

Marie's finding is interesting and Claudio's is not, but Marie was not curious (nor did she have any reason to be) and Claudio was (and had every reason to be). Now, on a Humean conception, Marie gained nothing of epistemic value because of her (appropriate) lack of curiosity. Yet we have the strong intuition that Marie feels the same glow that we receive when we satisfy curiosity in normal cases. It is not just that she has a new insight into economic theory, but she has one that is in some sense enjoyable. Her case is analogous to someone biting into what looks like mouldy bread and turns out to be both a nutritious and delicious cake. The opposite goes for Claudio. Now, some of this has to do with the difference between subjects and questions. Sometimes an inquiry into a boring question leads to an interesting answer—and vice versa—but it may *not* be the case that an inquiry into a boring subject sometimes leads to an interesting answer. The onus, however, is on the Humeans to show as much. I think there is more to be said on this matter, but will save it for later in the chapter.

Finally, consider what could be called the ignobility of Humeanism. It is insulting to a normal sense of the value of the pursuit of knowledge to say that all that its dignity is exhausted by practical instrumental usefulness and the satisfaction of curiosity. If that is all there is to it, then our epistemic pursuits seem to be disconnected from human flourishing or even from anything of value beyond preference satisfaction.

[113] This example is inspired by a similar one taken from Brady (280).

The worry might go deeper still. In what follows I make the case that if the Humean joins forces with the pragmatist she will at least be able to explain why the interesting is connected to human goods as taste is to nutrition; that is to say, the interesting is Virgil to life's Dante—not a perfectly reliable guide, but good enough to keep us more-or-less on a path worth following. Epistemic value is value in the service of something recognizably valuable, though it remains a mere means. But uncoupled from pragmatism, the epistemic life (qua epistemic life) is a spinning cog connected to nothing beyond a kind of pleasure—and even that connection has been, I hope, shown to be tenuous at best.

6.3: Groping Towards a Hybrid Account:

There is, however, something good to be said for curiosity-based accounts. They at least make some attempt to comprehend the numerically vast, and vastly differing ways in which we expend our mental energies. That is the obvious way in which one could be led into such a view: By observing that we really care about learning things that make almost no practical difference in our lives. This advantage mirrors the naive pragmatist's great disadvantage. If epistemic value is *purely* whatever helps us get to other things of value, then it is difficult to defend hobbies, curious interests, most of the humanities, and vast swathes of the sciences. In what follows I will first say a bit more about the disadvantages of pure pragmatism, then will turn to ways in which the two views can make up for (some of) each other's deficiencies.

Probably the most cited recent paper on the question on CA is Grimm's (2011) "What is Interesting?"[114] Though I think its conclusions are suspect, it is a serious work, and one that deserves serious consideration. In it, Grimm argues that the commonly held distinction between the practically valuable and the "purely" epistemically valuable is a fraught one; while not outright spurious, this distinction, in Grimm's eyes, is blurry. Fundamentally, the important or interesting is wrapped up with questions about how to live life, and as such are inextricably bound-up with practical concerns.

[114] The literature on this is miniscule. I have touched already on bits of it (e.g. Sider), much of which deals with the question of epistemically valuable content only tangentially. This Grimm paper is, surprisingly, the only one out there directly addressing these questions that has gained any traction.

In this chapter we will be building a broadly similar account that more directly addresses the relationship between the interesting and the useful.

I want to flag one complication from the outset. In his opening precis, Grimm says that, inter alia, he wants to "shed light on increasingly important and perplexing questions in the epistemological literature: e.g., questions concerning how to think about 'the epistemic point of view,' as well as questions concerning what is most worthy of our intellectual attention and why" (515). These questions are of great significance for my purposes, particularly the second one. The contemporary debate has, as I have said, spent too much energy on which epistemic states are fundamentally value-bearing, and not nearly enough on what subject-matter matters. But while Grimm admirably attacks the ignored questions, he does seem to simply assume that whatever is worth knowing "from the epistemic point of view" is *interesting*. This is a big assumption, and an unacknowledged one. For it may well be that much of what is epistemically important is boring to you or me—or perhaps even to the ideal observer.

In fairness to Grimm, he is breaking new ground, and we should avoid getting too caught up in fine distinctions; furthermore, he *does* deal with this question, albeit briefly, towards the end of his paper. Nonetheless, he repeatedly uses the language of "interesting or important" in the opening few sections of his paper, before gradually settling simply on the "interesting." This slide should not go unnoticed! It is easy to convince someone to be a Humean (or quasi-Humean, in his case) if they do not attend to the distinction between "questions it is important to answer" and "questions we like having answered."

With this worry in mind, we can begin. As mentioned above, Grimm is himself taking aim at what he thinks is an underanalyzed move. We all agree that knowing how to find one's dentist's office is of importance in the practical sense, and we all agree that it is *not* a very *epistemically pure* sort of importance. It matters for practical purposes, but, unless one's dentist works in Nirvana, it does not amount to a theoretical achievement worth mentioning. He will return to finesse this distinction in a way that partly preserves it—but not completely. For, as will become clear, he thinks the difference between the trivial but practically useful and the interesting is much less sharp than the intuition that

the dentist example is trying to provoke. Anything really interesting is interesting because of the way it relates to our distinctive human concerns about how to live our lives—fundamentally practical concerns. A good initial way of thinking about the distinction is one between distal and proximal usefulness. Knowing about dentists is proximally useful; knowing about protons is distally so. Or, here is a slightly more refined preliminary statement of Grimm's view: what is practically important is determined by the particular case, but what is epistemically important is set by paradigmatically human concerns on something like a typological basis. There are things we need to know as human beings, and we evolved in a way that is sensitive to those needs. When an individual diverges from the paradigm, so much the worse for her, for she will fail to appreciate what is interesting to humans.

Grimm's method for getting to this conclusion is a fitting one for this little-trammeled ground. He proposes what he takes to be two paradigm interesting subjects—the same two subjects that everyone else proposes as interesting:

1) Fundamental scientific laws

2) The existence of God (521).

In addition, he takes some paradigm cases of uninteresting claims with which the reader will be familiar (grains of sand, phone books, dentist's offices, etc.). He then identifies the questions "of epistemic interest, relative to which [(1) and (2)] could then be seen as instances—that is, basic topics that seem to underlie our interest in the topics that plausibly belong on the 'interesting' side of our ledger" (521-22). These fundamental questions are:

1) What is there?

2) How does it work?

3) How did it get to be that way?

These are extremely general questions that seem to cover most questions, full stop. This is no vice, for there are obviously very many interesting questions. The trick is to make sure uninteresting questions, which are infinite in number, are excluded.

In what follows I summarize his treatment of all three. I will argue that some of his claims about the relationships of priority among the three are mistaken, and that the tension between what is interesting and what people are interested in is basically unresolved. Nonetheless, it is as good a subjectivist analysis of the interesting as I have seen, and I don't supply a better way of filling in the details. Rather, I will critique his and then try to supply a general method for negotiating the tension between what is interesting and what we are interested in, and then supply a rule that resolves the Humean and pragmatic streams of subjectivism. That, not a detailed analysis of Grimm, is the goal of this chapter.

6.3.1: What is there?

The question of what there is seems to admit too much. Unless Plato is right, the blades of grass in your garden do exist! Grimm is aware of this problem. Roughly, he's looking to get at "structure" in the world, along the lines of Sider's "substantivity" in *Writing the Book of the World*. Says Grimm: "I think that the interest fueling the 'What is there?' question is best thought of not as an interest in cataloging every detail of the world but rather in figuring out what kinds of things there are in the world, or the sort of stuff that makes it up" (523).

He cites Sider approvingly:

> I think we get at a question that drives not just scientific inquiry but metaphysical inquiry as well. What both sorts of inquirers are interested in finding out, in their different ways, is what kinds of things constitute the world, or what it is made up of. Compare Ted Sider: 'The goal of inquiry is having one's belief state accurately reflect the world, which in addition to lack of error and lack of triviality requires one to think of the world in its terms. The ideal inquirer must therefore carve the world at its joints, otherwise her beliefs do not adequately conform to the world,' (Ted Sider, *Writing the Book of the World*, manuscript, 50)' (523).

What Grimm is trying to show is that the patch of sand in the Sahara is not interesting, for there are no integral "joints" in the world's body to be carved there. But he fails to show this. For, in the sense Sider is describing, there are joint carvings that are fairly trivial—including claims about patches of sand in the Sahara.[115] To see this, first consider questions that Sider claims are insubstantive. It's

[115] Grimm offers no account of "joint carving" of his own. His account is parasitic on Sider's.

admittedly a useful metric for cutting out lots of worthless questions. Is the pope a bachelor? Is some "green nonsense in a martini glass" a martini? On an intuitive construal of "substantive," such cases won't count. "Pope" and "bachelor" are at least minimally important concepts to have, but at their intersection lies a question that is useful only to philosophers trying to illustrate points. One candidate explanation for its insubstantivity is that the view that pope is a bachelor and the view that he isn't carve up nature equally well. Each candidate view is just as good as the other; nothing of any importance hangs on the matter. This is a good result, since we don't want to say that such questions matter. Unfortunately, it cuts out far too little. Lots and lots of questions are "substantive" in the sense that they do cut nature at its joints. *All* the trivial questions Grimm gives as examples are properly joint-carving in Sider's sense. The problem is not that they don't track the structure of reality (again, in Sider's sense), it's that they don't track very much of it. An arbitrary patch of sand has Siderian joints. "Sand" picks out something clear and real. Knowing about that patch of sand is indeed worthless (or close to worthless), just for other reasons.

I raise this—what looks like a rather meanspirited cavil—because I want to draw attention to the requirement that substantive or interesting or important questions get at the structure of the world, in Sider's sense. It's a good requirement, but it doesn't actually shed any light on what Grimm himself means to get at. Sider himself, in the above quotation, makes sure to add the requirement that the goal of inquiry require a "lack of triviality" *in addition* to being joint-carving. The structural substantivity requirement is meant to cut out questions about grue and bleen,[116] not to cut out inane inquiries into phone books or blades of grass. Something else must be doing that winnowing, and Grimm does not say much about what it might be. I want to conjecture that what he means to be getting at is, in a rough sense, hierarchical. God or the big bang don't merely explain "more" structure in the sense that they give us more true propositions that don't fail Sider's joint-carving test. (Recall: They might not even give us more propositions; counting propositions is vexing business.) Rather, they give us a

[116] See Goodman (1985) for a description of concepts that do not carve the world at the joints.

higher kind of structure. But everyone already knows the truth of *that* claim; it's just that we don't know exactly what said claim comes down to.

In what follows, Grimm points out that many interesting phenomena blur the joints of the world in a way that indicates a bit more about what he has in mind by "joints." Fireflies are living creatures that glow; water bugs walk on water; polar bears live on ice—most of the creatures nature documentaries depict are interesting because they don't fit what we commonly take to be the rules of the animal kingdom. Something is surely right about this. The more our epistemic map of the joints of nature comes under fire, the more interested we get. But here we see where our first worry starts to become a real problem. It's one thing to say learning about fireflies has value because fireflies have these interesting features that defy our normal assumptions about insects. It's quite another to say that knowing about fireflies is important, even from the epistemic point of view that concerns us. Fireflies are peripheral in roughly the same sense that phone books and Saharan real estate are peripheral. If their shine attracts the mind's eye away from deep matters, then some kind of epistemic cost in incurred. Since fireflies are *also* of little pragmatic importance, the interest we take in them is a problem for the sophisticated subjectivist.

6.3.2: How does it work?

Though the two bleed into each other, the question of how things work has, for Grimm, a kind of explanatory primacy over the question of what is there:

> A case can be made that we are interested in keeping track of joints is because we are interested in keeping track of differences in causal powers, and the reason we are interested in keeping track of differences in causal powers is because of our interest in grasping how the various parts of the world relate to one another… By contrast, taking the 'What is there?' questions as basic makes it a bit mysterious why we should be particularly interested in tracking joints in the first place, as opposed to counting grains of sand or what have you (527).

The joints are between things, roughly speaking. Part of the reason we want to know about certain kinds of things (those revealing "joints") is that we want to know how they work. Jointless bits of reality reveal no secrets of the functions and operations, governed by laws, that constitute the subject matter of our understanding of how things work. But we might ask: if we are interested in knowing

how things work, why do we need to care that much about what actually exists? Why not contemplate the workings of things that don't exist? Grimm's best recourse would be to argue that knowledge of counterfactuals (How would things work if they were different?) is indeed the sort of thing we're interested in, for counterfactuals reveal reality.

Many people find mythologies (Egyptian, Greek, Norse, etc.) deeply interesting; few believe in them. I think it fairly obvious that little of this interest can be reduced to keen curiosity about the historical peoples and places that generated these mythologies. The wrath of Achilles or the regeneration of Osiris are great stories—and more interesting to enthusiasts than, say, the anthropology of Asia Minor or the Nile delta. If Grimm is right about the priority of the functional question over the existential one, then this is no surprise at all. We want to know the workings of the politics of Mount Olympus because they explain the ancient Greek cosmos much better than most other things. I hasten to add that no one, Grimm and myself included, would want to say that *all* of the interest in these stories stems from the same source. Everyone would grant that great stories can be viscerally exciting in a very pleasing way.

I grant that it is also true that *some* counterfactuals are very helpful in understanding the actual world in various ways. First, if we assume (with most metaphysicians) that fatalism is false, knowledge about what *would* be true if the world were different in various ways is just that—knowledge. It's knowledge of reality, albeit a bit abstracted. Second, such knowledge is useful in all sorts of practical ways. If it's the case that most people are cowards to some degree, we probably ought not be too harsh on cowards; and reading Remarque or Hemingway is a good way (indeed, probably the only easily accessible way) to acquire evidence about whether they are cowards.

The pleasure of fictional immersion, the epistemic value of counterfactual knowledge, and the practical and ethical usefulness of counterfactual knowledge all form non-exclusive bases for the value of learning how things work. None seems to be exactly what Grimm is getting at—though I think all are in some way related to what he means. His focus, again, is on the interesting per se, which does not seem to reduce to any set of the above. Rather, the idea is that human beings have natures that drive them to understand the world in certain ways, and that the form that drive takes is

fascination. That we are interested in these things is what he is flagging; where the interestingness bottoms out is a separate question which he will try to answer once the system is built. For now, he simply wants to recognize the priority of the functional over the existential.

Nevertheless, the value of knowledge of fictional worlds either reduces to some combination of the above, or it doesn't. If it reduces to pleasure, we are left with pure Humeanism, which is beset by all sorts of problems. If it reduces to a kind of roundabout knowledge of the actual world, then we ought to be very careful about ascribing it any priority over existential ("What is there?") knowledge. If it bottoms out in a kind of practically useful skillset, then it looks like we may be leaving the realm of distinctively *epistemic* value. (This, I will argue, is nonetheless the Subjectivist's best bet.) Without a sense of distinctively epistemic value, it is difficult to see what epistemology really comes to as a discipline; further, it becomes much harder to make sense of why truth, say, is anything like an ultimate epistemic good. If the epistemic is just a handmaiden of the practical, then perhaps there is nothing epistemically out-of-order about believing salubrious falsehoods. And, if the value of knowing the fictional boils down to something else, I struggle to think what it might be.

This digression also makes it clearer why conflating the interesting and the important is so dangerous. *Of course* Greek mythology is interesting. The gods and heroes might be about as interesting (and probably more titillating) than whatever natural or supernatural forces actually govern the world. Is such knowledge as epistemically *important*? The intuitive answer is that it is not! But this subtle shift (from the interesting and important to the merely interesting) is not surprising given Grimm's *methodology*. If we *simply* go around looking at what most people find interesting (weird animals, Mount Olympus, etc.) then we are no better off than the ethicist who goes around looking at what almost all people *do* in order to figure out what is right. There is no built-in mechanism for taking epistemic bad judgment into account. Like me, you might think maps are endlessly interesting and can happily spend hours just staring at them; but if you were more epistemically virtuous you might focus more on e.g. the foundations of ontology.

None of this is to suggest that a better method would be to, say, take a survey of what people thought was epistemically important. Asking the man on the street what he thought was of pure

epistemic importance would by futile. Perhaps an ingenious empirical polling method could be devised that would not be so plagued by abstruse concepts and language, but I know of none. Nonetheless, if we share the intuition that some interesting things are unimportant, we can at least offer conjectures, as I have done, about what accounts for the difference.

6.3.3: How did it get to be that way?

We now turn to Grimm's third question. "The thought here is that with respect to almost any X, finding out 'the story of X,' or the causal history of X, is interesting to at least some extent: the story of the cotton gin, the story of Rutherford B. Hayes, and so on" (527). As Grimm admits, this question is not fully distinct from the other two. When we try to figure out how something got to be the way it is, we do so because we are interested in the causal contours of the subject; that is, "dependencies that hold between the current character of a thing and some aspects of the past" (528). Of course, talk of "cross-temporal dependencies" raises questions about whether the past exists, and, if it doesn't, whether there is enough of it somehow imbued in the present to ensure that, when we talk of "the past," that we are talking about *something*. If it is not real, is it worth investigating in the same sort of way that Middle Earth is worth investigating? If it is, is there anything *at all* distinct between the "how does it work" question and the "how did it get to be that way" question? I don't know the answers to these, but I note the lacunae because doing so reinforces previous chapters' suspicion that this sort of inquiry is only profitable when coupled with a metaphysical view.

6.3.4: A Fourth Question?

Grimm is a moderate Humean about epistemic value; such a view, if accepted, allows for a loosely empirical, psychological approach towards axiology, for it has fewer constraints. If human beings happen to find some unimportant-seeming-things interesting, the Humean need not worry. But, perhaps tellingly, Grimm *does* seem bothered by some common interests. Jennifer Anniston's love life, to take one of his examples, is a subject that sells a lot more magazines than Darwinism or Trinitarian theology. Now, if celebrity gossip is uninteresting, then it would be nice to have some kind

of error theory explaining why so many people spend so much of their time consuming it. Grimm doesn't examine any, but they are not difficult to generate. Here are two:

1) There is a pleasure in indulging fantasy. In our idle moments, we often find ourselves day dreaming about what it would be like to be rich or powerful or beautiful. We are less likely to day dream about being poor or impotent or ugly. This is because it is enjoyable to think about all the wonderful things we could do if we had all the things we want, so much so that we sometimes intentionally day-dream this way, even in the knowledge that the inevitable return to reality may leave us saddened.

2) Commonalities with idols seems ennobling. If we find out that Britney Spears is also a Sagittarius, or that David Beckham is also a Brexiteer, or that Barack Obama is also terrified of snakes (all fabricated examples), we feel as if we are not so different from our perceived betters. Perhaps they would even enjoy having a beer with us.

As with our discussion of mythology, none of these are exclusive. Also as before, all offer reason to be suspicious of explanations defending the properly *epistemic* value in knowing about celebrities. For, as before, Grimm argues that celebrity gossip *is* interesting, at least insofar as it provides a means of answering the question of most interest to us: "How should I live?"

This fourth question is certainly an important one. If its answers end up counting as being of *significant epistemic value*, as Grimm believes, then the distinction between the practically and epistemically valuable is hardly sharp. If the question "how to live well" counts as a purely (or perhaps "lastly") epistemic one, does it not hopelessly violate our intuitions that there is some real difference between practically and epistemically valuable knowledge? Some might be willing to simply jettison that intuition, but doing so is difficult. Grimm, though not insensitive to the worry, seems willing to fold *all* epistemic value into the realm of the practical. This is a promising strategy in a way, and one I will take up later in the chapter—but nonetheless a costly one. The three questions he began with, he suggests, might serve the fourth:

> It might be thought, moreover, that the final question bears no deep relationship to the other three… But there is a different story that one could tell about how these various topics are

> related, and one that seems to have a lot going for it: namely, a story on which our interest in these dependence relations is not intrinsic—not something we are interested in 'for its own sake'—but rather derives from our more basic practical interest in prediction and control (533).

He goes on to lay out the view a bit more thoroughly, but the broad picture is already clear. Questions about what is, and how things work, and how they got to be that way reveal dependencies of various sort: causal, mereological, temporal, etc. These in turn allow us to "manipulate and control:"

> They tell us how, if we were to change the value of one or more variables, we could change the value of the other variables… If these are relationships that are potentially exploitable for purposes of manipulation, there is no mystery about why we should care about them" (Woodward 2003: 6-7, quoted in Grimm 533).

One advantage of this view, beyond its laudable simplicity, is that it implicitly addresses one of my earlier objections. For ease of memory, I reproduce the claim to which I objected:

> A case can be made that we are interested in keeping track of joints is because we are interested in keeping track of differences in causal powers, and the reason we are interested in keeping tack of differences in causal powers is because of our interest in grasping how the various parts of the world relate to one another… By contrast, taking the 'What is there?' questions as basic makes it a bit mysterious why we should be particularly interested in tracking joints in the first place, as opposed to counting grains of sand or what have you (527).

In response, I made the case that a foundational interest in how things work need not presuppose all that much interest in what things (or, more properly, what kinds of things) exist. We can figure out how all sorts of systems work without knowing whether their constitutive parts exist. This, however, seems to be a mark against Grimm's view. (Namely, that it will lead to the conclusion that knowledge of far-flung fictions like the Star Trek universe, which does not exist, is likely to be more valuable than knowledge of the world.) We can now imagine a rejoinder along the following lines:

> The interesting is best understood as instrumentally valuable. We find ourselves drawn to systems because figuring them out helps us navigate the world; however, as systems become more distant from reality, they are commensurately less help in navigating the actual world, and thereby cease to serve their purpose. As such, the question of what exists has a kind of priority over the question of how things work (and how things came to be the way they are), though only insofar as existential questions tether us to our practical aims.

This may or may not be a revision of Grimm's actual view, which does not go into such details, but if it is, it is not a radical one. At the heart lies his concern with the question of how to live a good life; this amendment just foregrounds that aim. As a final caveat, I should add that the functional questions (how things work and how they got to where they are) retain some situational priority over existential ones (what exists), for it is indeed the case that knowing about the "joints" of reality is of greater instrumental utility than many mundane existential questions.

Grimm's view is Humean in a loose sense because at its core, epistemic value is a function of our own practical purposes.[117] But it is not of the crude sort whereby the mere satisfaction of curiosity and pleasures therein exhaust epistemic value. Nonetheless, there is little to distinguish the epistemic and practical realms. I now turn to two distinctions that facilitate understanding and discussion of the matter.

6.4: Two Helpful Distinctions:

First, I want to distinguish survival and flourishing. I use these terms rather loosely, and intend no controversial claims about biology or the philosophy of biology. I do think (and will take for granted) that, broadly speaking, evolution selects for survival and reproduction, but that evolutionary success is not identical to human flourishing on either the individual or the species-wide scale. There is overlap, but not perfect overlap. Regarding survival, I have nothing interesting to add to debates about whether the fundamental Darwinian unit is the gene, the individual, or the species. I am using "flourishing" especially loosely. I mean to include views wherein flourishing is a matter purely of getting what we want, as well as views wherein flourishing is to any degree independent of our aims.

Second, I want to mark the difference between retrospective and prospective fitness. With respect to either flourishing or survival, traits can either be helpful in the future or they can *have been*

[117] This may be so anyway. It may be better understood as pragmatist in the sense I outlined at the beginning of the chapter. Either well, not much of importance hangs on this exegetical question.

the sorts of traits that *were* helpful in the past. Environments are fluid, and traits that were selected for in the past are sometimes less helpful in the future, and vice versa.

We can map out the logical space in the following simple table:

	Retrospective fitness	Prospective fitness
Flourishing promoting	Doxastic states that in the past promoted flourishing	Doxastic states that will promote flourishing
Survival promoting	Doxastic states that in the past promoted survival	Doxastic states that will promote survival

Which of these are of use to Grimm and other subjectivists? In what follows, I will argue that a "prospective flourishing" view is the best bet, though none of them are without complications.

The two survival-conducive options are plagued by worries about the gap between mere Darwinian success and what almost everyone believes to be the sorts of successes at which we ought to aim. There are abundant examples of kinds of curiosity that, while plausibly the product of crude evolutionary design, either don't help us as much as other kinds, or actively undermine our ends. Grimm's own example of celebrity gossip may be the by-product of a kind of tribalism that, while once necessary, is now outdated. Plenty of celibate people have happy, productive lives in part because of their celibacy; for them, curiosity about kinds of relationships or sex may serve as an impediment to flourishing. Furthermore, it is important to remember that we are *not* perfectly programmed to survive in the first place. Evolution is clever, but filled with dysgenic mutation and imperfection. So not only do these views sometimes deviate from the kind of flourishing we think of as paradigmatic of success, they do not even pick out what we find interesting with high accuracy.

The two flourishing-conducive views have an impediment similar to the second worry above. Lots of truths that lead to flourishing are magnets for our curiosity—but not all. Many are extremely dull. The problem is that it is not clear where this leaves the status of curiosity or interestingness. Curiosity can be a useful horse upon which human flourishing can ride, but it is a stubborn and

sometimes quite wayward one. The connection between epistemic value and the interesting or curious is, on this picture, a loose and contingent one, where curiosity does not underlie epistemic value, though it does sometimes help us get to it.

The distinction between the prospective and retrospective views echoes the above distinction somewhat. Roughly speaking, prospective fitness is analogous to what is flourishing conducive, and retrospective fitness is analogous to what is survival conducive. The difficulties with each side of this second distinction nonetheless need a little spelling out.

Retrospective approaches lash epistemic value to those traits which have been instilled by evolution in the past. Their strength is that evolution, if construed very broadly, can account for many of the traits, epistemic and otherwise, that we actually have. Retrospective views are views about what is already there, as it were, within the agents in question. Prospective approaches are based in what traits *should* be present in those agents, given the spatiotemporal landscape in front of them.[118] Environments change, and it takes time for evolution to catch up. Prospective views are those that hold that the norms and axiology governing epistemic value are a function of those epistemic properties that will further the agents' survival or flourishing *now*. The advantage prospective views have is that they mandate norms or axiologies that further intuitively appealing goals, be they survival or flourishing.

There are complementary disadvantages to each view. Retrospective approaches fail to keep up with our environments. They will dignify traits (e.g. curiosity) and outcomes (e.g. belief) that are in keeping with whatever wisdom evolution has managed to preserve in the present environment, but will *also* dignify traits and outcomes that are no longer beneficial. Like survival views, retrospective views hold as valuable some attributes that hurt us, and undershoot some that help us. For instance, it

[118] I have no view regarding how fine-grained the "future" in question should be. We can imagine the exact proclivities that will aid some given agent in her exact environment, or we can imagine something more general, that does not take into account absurdly precise environmental details. Suppose some agent, Charles, will win a great sum of money on a quiz show in thirty years if he knows enough about Leopold of Belgium; as such, curiosity about Belgian royalty may count as conducive to his survival or flourishing. We can perhaps imagine such curiosity as flourishing-conducive in Charles' case, or we can prefer a more general account that doesn't take into account such extremely contingent futures. However, I don't think much I say hangs on which account is better.

may be the case that we have evolved to be very curious about politics, because in the past our societies were much smaller and less politically inert. Now, we remain very curious about politics, despite the fact that (plausibly) almost all of us are wasting our intellectual energies on questions over which we have no control. Prospective approaches, on the other hand, risk abandoning curiosity or "the interesting" altogether, and collapsing into reductive kinds of pragmatism. While they elevate what actually helps us now, they call valuable that which we do not find interesting.

All these worries can be categorized as either failures to accurately *correspond* to agents' interests and curiosities or failures to set *goals* worthy of the name. Retrospective and survival views do better at accurately capturing what we actually *find interesting*. They better fit the Humean intuition that our interests matter. Prospective and flourishing views do better at ensuring that epistemic accomplishments actually *help us live good lives* (or, perhaps, they partially constitute good lives). The question, then, is whether we should favor proper correspondence or proper goals.

My view is that goals are more important, for reasons that I have already touched upon. First, correspondence views fail to correspond well. What they hold is that some epistemic achievement is valuable insofar as it is the product of the Darwinian inheritance that we now possess. This will capture a great deal of our curiosities, but it will not capture them all to the correct degree. Crucially, it simply is not the case that *any* of these views will achieve perfect correspondence with our curiosity or interests. The only way to ensure that our actual epistemic desires are always worth gratifying is *directly*, through a kind of naïve, dogmatic Humeanism. A sort of "gut-check" shows how strange the retrospective and survival-conducive views are. If we are wondering how much epistemic value we have garnered in a particular inquiry, we must first answer to what degree our curiosity is a product of past evolution (in the case of the retrospective view), and second to what degree our curiosity fits the goals of survival and reproduction. It is nearly certain that answers will often be "not at all" or "very little," even for matters we find very interesting. Now, admittedly, these positions get a lot closer to capturing our actual interests than their counterparts. But they fail to satisfy the Humean yearning for our interests *as our interests*. No committed Humean should be happy with such a pyrrhic victory.

None of the proposals, as they presently stand, do particularly well with the Humean intuition. But the prospective and flourishing-conducive proposals do very well indeed with the Pragmatist intuition. That which will help you do well (and not just survive), and that which will help you do well in the future (and not just that which would have helped you do well up until now). The pragmatist intuition more-or-less just is the view that the basic epistemic value is that which, upon our possession of it, promotes our flourishing, or some component of our flourishing. The cost of the view is that it excludes not just wayward curiosities that snuck in under evolution's nose, but many curiosities with proper pedigrees.

Grimm's pioneering paper is concerned with breaking ground, not negotiating fine distinctions like these. Nonetheless, at least one of the four views I detail above has to be plausible in order for his hybrid-style Subjectivism to gain any traction. He does address some of these concerns implicitly, but does not go quite far enough. I will comment on his implicit responses, and then offer a fuller defense of my favored Pragmatism-friendly view. While I do not find it satisfactory, I think it goes some of the way towards making subjectivism about epistemic value viable.

6.5: A Final Word on Grimm:

Grimm is awake to some of the worries above. He gives the example of his young son's interest in the solar eclipses. Solar eclipses are fairly unimportant for human flourishing today; understanding them does not help us achieve our real or ideal aims very much—at least not to the extent that would account for his son's deep fascination. Curiosity about them is real, but fails to be prospectively useful. Whether it is flourishing-conducive, or even survival-conducive, is also dubious.

Grimm seems to favor a retrospective survival-conducive response. For many centuries, some thought eclipses portended great events (e.g. God's wrath). To know about their real workings then would have been of great practical help. This in turn provokes the acknowledgement that evolution has forged us into epistemic generalists. Since there is no way to design a species such that it always and everywhere seeks out only those systems (or phenomena, causal history, and so on) which are of

practical utility, the best we can hope for is a kind of roving curiosity about the *kinds* of things which are likely to be of help. Says Grimm:

> On this view, we should not suppose that the interest of things outside of our control is in some sense intrinsic; instead, it is best thought of as a kind of offshoot or hangover of our original instrumental interest, now applied outside its original domain (536).

As I have already said, this is a costly move. A riff on an earlier analogy nicely enfleshes those costs: Just as there is something defective about eating a sugary dessert when one has healthier alternatives available, so there is something defective about indulging one's desire to learn a lot about eclipses when one can focus on something more substantive.

I am here bracketing many important questions here. Does the axiological standard need to be species-derived? Why not on the individual scale? Are some inquiries actively harmful—akin to eating poison—or do our natures give us such wide latitude that *every* inquiry falls under the description "instrumental for some purpose *or* a "hangover" from some earlier purpose?" A thorough answer would probably require some kind of precisification of his view. I am happy to settle for a less-than-thorough critique, because that seems much more realistically achieved.

In any case, the worry I have is that this sounds cruelly reductive. Nonetheless, on Grimm's picture, investigating large portions of physics (those that offer no practical advantages) really is akin to stuffing oneself with nutritionally-deficient candy. This, even by itself, seems a dim view of inquiries most of us see as noble. But it will also imply that any bit of knowledge that is *uninteresting* to our still-simian, somewhat backwards minds will count as epistemically valueless, as a kind of normative trespassing or (more cautiously) as at least *relatively* deficient. I take it as obvious that interestingness comes in degrees, so it's possible to devote one's time to relatively less interesting matters. This, I think, betrays a very high view of human nature. The questions Grimm himself takes to be paradigmatic (the existence of God and the fundamental laws of physics) seem not to be as interesting to most people as celebrity gossip. This is consistent with the claim that people would find those "deep" questions much more interesting if they were better able to understand them. What is interesting, for Grimm, is (roughly) what we find interesting, though he does provide a few caveats

and qualifications. But since this is largely true on his account, it is beyond doubt that some "low" inquiries of the sort undertaken by tabloids will count as more valuable (or maybe obligatory, healthy, norm-conforming etc.) than some "high" inquiries undertaken by academics.

Suppose then that we should not understand his account in terms of axiology-generating or normativity-generating teleology. This seems too weak. If some rogue evolutionary impulse, unprepared for the information age, directs us to learn all there is to know about some superhero franchise, and if that impulse itself is a product of our practical concerns, have we not in some way failed to get to the heart of "important or interesting" knowledge? But, more importantly, a non-teleological (at least non-normative) account seems to whittle down the central thesis to a few basic claims that most everyone already accepts. *Of course* evolution selected some of our epistemic faculties to be adaptive. *Of course* that will affect what we want to know about. *Of course* this desire will be coarse-grained and will fail to be applied adaptively sometimes. But, I take it, this isn't all that controversial. It might, admittedly, be a bit more controversial to claim that this short litany tells us more-or-less everything we need to know about the interesting. But this is a fairly sweeping reduction, and thus incurs an argumentative burden. Finally, while this account may have value as a psychological or anthropological inquiry, it will tell us nothing about epistemic value. That may not be a failure on Grimm's part, but it is certainly bad news that the most cited paper on content-based epistemic value turns out to say very little about epistemic value.

6.6: A Detour Back into Moral Responsibility:

I close the chapter by rebuilding the subjectivist view on more explicitly pragmatist lines. The result will not please everyone, but it carves out a place for Humeanism that may mollify Humeans to some degree. Moreover, it avoids many of the stranger conclusions Grimm's view reaches.

Historically, many analytic philosophers defended a consequentialist conception of moral responsibility norms.[119] The purpose of such accounts is to provide not just a schema detailing how

[119] See Vargas (2008, 117) for a survey. In particular, see Smart (1961).

the norms of praise and blame work, but a defense of their use. For instance, one might think that ascriptions of blame are justified because they have a deterrent effect on future bad behavior.

The view has fallen out of favor for many reasons detailed in Vargas (2008). Too many bedrock norms of moral responsibility do not seem justified on consequentialist grounds. How can we blame the dead? And why not refrain from holding responsible those who, though innocent, would be improved by negative reactive attitudes like blame? Praise and blame seem to be backward-looking, but consequentialist justification is forward looking.

Vargas himself thinks this consequentialist approach has been unfairly dismissed. I am not here interested in whether he is right, but rather in the structure of his defense. He divides the theory of moral responsibility into four subtheories: one for *agents*, one for *norms*, one for *the justification of norms*, and one for *aims*. The theory of agency tells us which sorts of people are apt targets for moral responsibility. The theory of norms tells us the content of justified responsibility norms. The theory of justification tells us why those norms are justified. The theory of aims is "an account of what the system of norms, practices, attitudes, judgments, and concepts associated with responsibility is properly understood as directed at" (97).

Our interest is primarily in the third category, the justification of norms. Vargas' thesis is that once we conceive of moral responsibility as a two-tiered system, many of the problems with traditional consequentialist conceptions of it go away. He assumes "(1) that the aim of the responsibility system is to foster a kind of moral considerations-responsive agency, (2) that the correct account of *responsible agency* is one that broadly favors more considerations-responsive agency and (3) that a moral influence account is the correct account of the *justification* of the norms of praise and blame" (99). With these assumptions in place, he summarizes the goal of his article thus: "[What] we get is the following account of the justification of the norms of moral responsibility: we are justified deploying responsibility-characteristic practices where such practices, as a whole and over time, aid responsible agents to act in ways governed by moral considerations. Appropriately holding an agent responsible involves rightly regarding them as a responsible agent and correctly applying the justified

norms of praise and blame, norms that derive their justification from their collective effects on fostering responsible agency" (99).

A similar structure is very helpful for the Humean Pragmatist. (Compare: "We are justified deploying epistemic practices where such practices, as a whole and over time, aid epistemic agents to garner useful beliefs. Correctly enquiring involves rightly regarding the targeted knowledge as worthy or interesting and correctly applying good epistemic practice—norms that derive their justification from their collective effects on fostering useful beliefs.") Prospective flourishing-conducive pragmatism is the goal, but a Humean-friendly rule is the engine that gets us there. Thus, even if it is the case that some particular tokening of curiosity-satisfaction is not unhelpful, or even hurtful, the system itself is better off for having a pro-Humean component than it would be otherwise.

6.7: Values and Rules:

The concern of this chapter, like the previous two, is epistemic value. I intended to confine the enquiry to axiology when I set out to write it. Unfortunately, I have found no way to accommodate the Humean intuition without straying away from value and towards norms. It isn't obvious how to relate values and norms in such approaches. I have already said a little on the matter, but it's worth clarifying further in order to see the limitations of such views. If we examine two-tiered approaches in general, they seem invariably to prescribe *rules* because following those rules tends to produce more or better *goods* than we would get by trying to acquire the goods directly. Rule utilitarianism, for one, claims that following certain rules leads to better consequences than any other strategy. To borrow an example from Vargas, following the rules of a sport makes the sport more enjoyable in the aggregate, even if individual cases of rule-following will sometimes be tedious. Curiosity-preference-satisfaction can be turned into a rule of sorts. One might insist on always pursuing what one finds interesting. Weaker versions are easy to formulate, in the form of qualified scopes ("Always follow your curiosity in such-and-such circumstances.") or weaker obligations ("There is a pro-tanto reason to follow your curiosity, and you ought to do so unless something else overrides."). These weaker rules are more plausible, as I will argue momentarily.

First, however, something needs to be said about the difference having a *rule* makes. We can make a distinction between rules and strategies and general policies, where rules (however limited their scope may be) are exceptionless, and strategies, say, are not. This is a non-trivial distinction, because whether this hybrid view really counts as Humean depends on it. First, the distinction.

The difference between "rules" and "strategies" as I am using the terms can be illustrated easily. Suppose I want to lose weight, and that I dislike diet and exercise. Suppose further that of the various weight-loss strategies, I dislike tennis the least. I might form a strategy to use my relative affection for tennis to lose weight by playing tennis frequently, but be open to changing and supplementing that strategy as needed. Perhaps I discover a diet or sport I like better, or start to sour on tennis because I am playing too much of it. I can adjust the levels as prudence demands. Suppose, conversely, that I know myself well enough to know that I need an inflexible rule ("Play tennis four times a week.") because otherwise I will rationalize myself into doing insufficient exercise. I might make a rule that is either strong or weak. Perhaps my rule excludes doing other forms of exercise, for they will sour me too much on the whole endeavor. Perhaps, conversely, I have a weaker rule, where I am allowed to skip tennis if my friend, whom I like spending time with, offers to lift weights with me that day. What distinguishes rules and strategies is their flexibility with respect to the goal, in this case the goal of losing weight. So, as I am using the terms, the fundamental distinction between rules and policies concerns cases in which an individual deviation will help bring about the goal. If the deviation is licit, one is following a strategy. If it is not, one is following a rule[120]. This is not an unfamiliar distinction. It is the same distinction between act and rule utilitarianism. But it is an important one here.

Returning to the subject at hand, we might wonder whether an exceptionless rule is appropriately applied to epistemic axiology. A rule of the sort might look like any of the following:

[120] This is true of very weak rules as well, so long as they are not so weak that they are mere tie-breakers. For example, one could have a rule that only favors φ-ing over doing something else so long as all options bring about exactly the same amount of the goods at hand. I am happy to call this a rule, even though strictly-speaking it fails the test above. That is, there are no individual tokenings in which one follows the rule and thereby loses out on some of the goods that the rule is designed to maximize.

1) One ought always pursue more interesting subjects (or curiosity-satisfying-knowledge) over less interesting subjects (or less curiosity-satisfying, etc.)

2) One ought to pursue more interesting subjects over less interesting ones once one has achieved a certain level of well-being, i.e. once enough of ones quotidian and professional needs have been met.

3) One ought to pursue more interesting subjects over less interesting ones to a moderate degree; that is, a degree that can be outweighed by sufficiently strong instrumental value.

4) One ought to pursue more interesting subjects over less interesting ones all else equal.

Suppose no such rule is appropriate. It will still very likely be the case that interest and curiosity play roles in epistemic strategy-formation, as I am using "strategy." It is obvious that one should take into account one's interests when pursuing various kinds of research, because our interest enables us to do more epistemic work than we otherwise could, and second because the joy we receive in pursuing our interests is itself a worthy practical aim. If I am pursuing financial success, find computers and mathematics very dull, but find other people extremely interesting, it would be better for me to learn how to be a good salesman than to be a good programmer because I have limited discipline and the latter would be so much of a chore that I would be likely to fail. Further, it is better to learn something enjoyable than to do a chore, independently of the consequences. And, finally, it might be the case that our interests have been at least moderately well-designed by the semi-visible hand of evolution to lead to our flourishing, and we ought to trust our interests at least some, even when we cannot see the benefits fully.

Such an approach really does not deserve to be called a "two-tier" or "hybrid" view. This in itself is no disadvantage. Indeed, it might count as an advantage, because it will not be plagued by well-known objections of the sort usually leveled at Rule Utilitarianism (Namely, that it is either a confused sort of Act Utilitarianism, or else a worse version of Act Utilitarianism that occasionally recommends the wrong acts.). But it is certainly more reductionistic. It concedes very little to the Humean, who, for all her faults, does, in the way I suggest, have a way to conceive of epistemic

values that doesn't end up more-or-less nullifying what has long been a serious philosophical question. If epistemic value is no more distinct from instrumental value than is monetary value—if it is just one more way to get what we want—then some serious revision of the state of philosophy is necessary.

Both Humeanism and pragmatism are subjectivist views insofar as the place the source of epistemic value within the agent rather than in the world. The strong version of the hybrid, the "rule" version has the advantage of generating a very plausible epistemic norm like the ones above, in a way that makes sense of our intuition that sussing out the interesting parts of the world has a kind of dignity and priority of its own, independent of our practical aims. It lends license to interests and hobbies, and can be the foundation (albeit a very partial one) to epistemology itself. And it does all this while respecting the fundamental subjectivist intuition that epistemic value must in some sense be tied to the agents for whom that value is valuable. It can even partially rescue our intuitions about the value of what I have been calling deep truths, at least insofar as those deep truths tend to be interesting ones.

Earlier I discussed the difficulty subjectivists have in accounting for how interested people are in fireflies. What I provide here is a partial solution for them—but only partial. Fireflies are interesting, presumably, because our psyches have been crafted in such a way as to find them interesting. Perhaps, as Grimm suggests, we are interested in them because they violate the normal rules about bioluminescence, and in a world where only one insect was *not* bioluminescent, we would find *that insect* interesting.

What I have here proposed is to establish a two-level view according to which epistemic value bottoms out in the value of practical utility, but supplemented by a rule mandating that we—at least to some degree—defer to our interests because evolution is wiser than we are. By doing so we drive forward the sciences and useful scholarship better than we would by simply focusing on the useful. The invisible wisdom of evolution, the way doing what we enjoy forestalls burnout, and the bonus enjoyment we get along the way together plausibly justify the rule.

Importantly, however, the rule in question is just a *rule*. It provides a justified decision-procedure (at least if two-tier systems are themselves coherent and justifiable). It does *not* give subjects like fireflies—*a fortiori* subjects like the Star Trek universe—epistemic dignity any more than rule utilitarianism makes the outcome of every specific rule-application better than any alternative. It is therefore *not* the case that e.g. learning about fireflies is epistemically valuable.[121] Our solution thus only goes so far in accommodating Humean intuitions that our epistemic hobbies are valuable. It should pacify the Humean only to the degree that rule utilitarianism pacifies the Kantian deontologist.

This is a major worry for the theory for reasons not solely dependent on the plausibility of Humeanism. So-called "deep" questions (of the sort we tailor Objectivist views to) are usually interesting questions to those with the capacities to grapple with them. I do not mean to imply that everyone with a capacity for metaphysics or quantum physics finds them interesting, but they are clearly interesting enough to large bodies of people. Humeanism can smuggle a lot of these truths in through the back door. The hybrid view, at best, can only mandate that those who are curious pursue them *without* acknowledging our intuitions about their presumptive final value.

If the problem of final value can be overcome, one solution might be to add in a second rule, akin to the Humean-friendly rule, that mandates some deference to what objectivists find important. But I think such a rule might fail to be compatible with the pragmatist foundation of the project, for two reasons. First, it is not at all obvious that the answers to deep questions are worth the effort from a practical standpoint, even from a two-tiered practical standpoint. Second, deep questions sustain our intellectual efforts (and thereby gain a credit on the pragmatist ledger) only insofar as they are interesting; however, anything interesting is already "covered" by the Humean norm. Anything else added will be, by subtraction, uninteresting.

Nor have we rescued epistemology from pragmatism. The value of truth is contingent on this view, for some falsehoods are likely good to believe, at least practically speaking. The distinctiveness

[121] Nothing hangs on the example. It is possible that learning more about fireflies is enormously practically useful. If so, then a myriad of other examples of things people find interesting will not.

of epistemology as a discipline will have turned out to be mostly illusory—a big revision indeed. As such, I remain skeptical of the subjectivist project.

6.8: Conclusion:

In the previous chapter I laid the groundwork for a more developed objectivism. In this chapter I have attempted the same for subjectivism. I do not find the view persuasive, but subjectivism undoubtedly has its charms. In the end, a much deeper exploration into true reductivism seems motivated. Having reflected on Humeanism and its implications, we can now see it as something of a side-show. Attempts to defend it, I hope to have shown, do best when subordinating it to broader reductive pragmatism. The battle for subjectivists, going forward, is with mainstream epistemology, in particular with those epistemic consequentialists who hold that the value of truth is great and irreducible. I do not think such a fight will be easily won, for the simple reason that jettisoning foundational, irreducible epistemic value subordinates truth itself to our needs and desires, and philosophers (rightly, I hazard) hold truth in high esteem. Nonetheless, the view deserves more attention than it has received.

Concluding Remarks:

I first argued that Velleman-style cases, (those in which the doxastic state of the agent has an effect on the truth value of the proposition held) have far stranger consequences than previously noticed. This is at once a reason further to explore such cases and a reason to ignore them. The reason to explore them is obvious. If they falsify widely held views in epistemology, then we ought to know as much. The reason to ignore them is the Moorean reverse: if these cases have deeply strange consequences for independently plausible principles, then we have some reason to think either that something has gone wrong in the construction of the cases, or that their application is so narrow that they merit at best a footnote to previously held principles. My suspicion is that the footnote view is correct, but I hope to have motivated further exploration on the reasonable chance that I am wrong.

The second chapter was defensive. There, I attempted replies to two unanswered critiques of volitionism. The first held that, because the concept of discretion lies at the heart of volitionism, and because we lack discretion over our intentions (for which we are clearly morally responsible), volitionism is false. I replied that volitionists have something other than discretion at the heart of their account. The second unanswered critique held that moral responsibility ought to be understood negatively, as that which is present in the absence of e.g. duress or compulsion. I argued that the central *reductio* in that argument applied to the negative account just as well, and that the argument relies too heavily on mere verbal considerations anyway. If I am correct, then philosophers of moral responsibility ought to stick to the central debates over tracing culpability, and should not be distracted by the "end run" arguments I criticize, despite their notable subtlety and creativity.

My hope is that the third and fourth chapters serve as motivation for a shift. There I argue that three of the greatest minds in the history of philosophy pay serious attention to the question of content-based epistemic value, which is some evidence that we, also, should pay attention. Historians of philosophy will certainly have more and better things to say on how Plato, Aquinas, and Hume approach CA, but *that* they approach it seriously is not in doubt. In the subsequent chapter I look at how CA has been approached in the much more popular debates about SA, and conclude that very

little has been accomplished through this oblique approach. A frontal assault may, therefore, be motivated.

In the fifth chapter I attempt to lay the groundwork for such an assault, particularly the objectivist one to which I am attracted. Attention to metaphysics and psychology—particularly the former—exposes potential fault-lines and offers a method for interrogating the strong intuitive differences likely to be seen among philosophers who take CA seriously, be they hedgehogs or foxes. The contours of our view of the world will shape the silhouettes of what we think worth investigating. Likely, then, this sort of question must be approached within philosophical schools, not across them.

The final chapter is both an invitation and a threat. The invitation is for those with Humean and reductivist sympathies to see how their views can be harmonized. Reductive pragmatism is not as strict as we might have thought, pre-theoretically, and is capacious enough to allow for serious Humean structure on a separate level. However, this reconciliation can only go one way, and that is with pragmatism at the helm. If this is right, and the value of truth can be reduced to the instrumental, then there is good reason for philosophers to worry about where this leaves epistemology.

Bibliography:

Ahlstrom-Vij, K. and Grimm, S.R. (2013), 'Getting it right', *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 166(2), pp. 329–347.

Alighieri, D. (2008), *Paradiso*. Reprint edition. Translated by R. Hollander and J. Hollander. Anchor Books.

Alston, W.P. (1985), 'Concepts of Epistemic Justification', *The Monist*, 68(1), pp. 57–89.

--- (1989), *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, New York: Cornell University Press.

--- (2005), *Beyond 'Justification': Dimensions of Epistemic Evaluation*. Ithaca, New York: Cornell University Press.

Annas, J. (1981), *An introduction to Plato's Republic*. New York: Oxford University Press.

Anscombe, G.E.M. (1963), *Intention*. 2nd ed. Oxford: Blackwell.

Aquinas, T. (1911), *Summa Theologica*. Translated by the Fathers of the English Dominican Province. London: Burns, Oates, and Washbourne.

--- (1952), *Quaestiones disputatae De veritate* (3 vol). Translated by R.W. Mulligan, J.V. McGlynn, and R.W. Schmidt. Chicago: Henry Regnery Company.

--- (1975), *Summa Contra Gentiles*. Translated by A. Pegis et al. London: University of Notre Dame Press.

Archard, D. (2007), 'The Wrong of Rape', *The Philosophical Quarterly*, 57(228), pp. 374–393.

Audi, R. (1993), *Action, Intention, and Reason*. Ithaca, New York: Cornell University Press.

Baltzly, D. (1997), 'Knowledge and Belief in Republic V', 79(3), pp. 239–272.

Becker, S. (2016), 'Resolving Turri's Puzzle about Withholding', *Dialectica*, 70(2), pp. 229–243.

Berlin, I. (1998), 'The Hedgehog and the Fox: An Essay on Tolstoy's View of History', in H. Hardy and R. Hausheer (eds), *The Proper Study Of Mankind: An Anthology of Essays*. London: Random House, pp. 436–498.

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

Brady, M.S. (2009), 'Curiosity and the Value of Truth', in A. Haddock, A. Millar, and D. Pritchard (eds), *Epistemic Value*. Oxford University Press, p. 0.

Bremer, M. (2008), 'Do Cats Have Beliefs?', in S. Hales (ed.), *What Philosophy Teaches You About Your Cat*.

Chisholm, R.M. (1989), *Theory of Knowledge*. 3rd edn. Upper Saddle, New Jersey: Prentice Hall.

Clarke, R. (2008), 'Autonomous reasons for intending', *Australasian Journal of Philosophy*, 86(2), pp. 191–212.

Clifford, W.K. (1877), 'The Ethics of Belief', in *The Ethics of Belief and Other Essays.* Amherst, New York: Prometheus Books, pp. 70–97.

Conee, E. *et al.* (2004), *Evidentialism: Essays in Epistemology*. New York: Oxford University Press.

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

--- (1978), 'Intending', in Y. Yovel (ed.), *Philosophy of History and Action*. Dordrecht: D. Reidel.

De Haan, D. (2015), 'Harmonizing Faith and Knowledge of God's Existence in St. Thomas', in H. Goris, L. Hendriks, and H.J.M. Schoot (eds), *Faith, Hope and Love. Thomas Aquinas on Living by the Theological Virtues*. Leuven: Peeters, pp. 137–160.

DePaul, M.R. (2001), 'Value Monism in Epistemology', in M. Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. New York: Oxford University Press.

Descartes, R. (1967), *The Philosophical Works of Descartes* (2 vol). Translated by E.S. Haldane and G.R.T. Ross. Cambridge: Cambridge University Press.

Evans, G. (1982), *The Varieties of Reference*. Edited by J.H. McDowell. Oxford: Oxford University Press.

Feldman R. (2002) 'Epistemological Duties,' in Paul K. Moser (ed.), The Oxford Handbook of Epistemology (OUP).

Fine, G. (1978), 'Knowledge and Belief in Republic V', 60(2), pp. 121–139.

--- (1990), 'Knowledge and Belief in Republic V-VII', in S. Everson (ed.), *Epistemology*. Cambridge: Cambridge University Press.

Fischer, J.M. and Tognazzini, N.A., (2009) 'The truth about tracing,' *Noûs*, 43(3), pp.531-556.

Frankfurt, H.G. (1969), 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy*, 66(23), p. 829.

Friedman, J. (2013), 'Suspended judgment', *Philosophical Studies*, 162(2), pp. 165–181.

Fritz, K. (2018), 'Moral Responsibility, Voluntary Control, and Intentional Action', *Philosophia*, 46, pp. 831–855.

Gilson, E. (1938), *Reason and revelation in the middle ages*. Edited by J.K. Farge and W.J. Courtenay. New York: C. Scribner's Sons.

Goldman, A.I. (1999), *Knowledge in a social world*. New York: Ox.

--- (2004), *Pathways to Knowledge: Private and Public*. Oxford, New York: Oxford University Press.

Grimm, S. (2011), 'What Is Interesting?', *Logos and Episteme*, 2(4), pp. 515–542.

--- (2008), 'Epistemic Goals and Epistemic Values', *Philosophy and Phenomenological Research*, 77(3), pp. 725–744.

Haddock, A., Millar, A. and Pritchard, D. (eds) (2009), *Epistemic Value*. Oxford, New York: Oxford University Press.

Hare, R.M. (1981), *Moral Thinking: Its Levels, Method, and Point*. Oxford: Oxford University Press.

Hawthorne, J. (2005) 'The Case for Closure', in *Contemporary Debates in Epistemology* Blackwell: Malden, MA.

Hempel, C.G. (1965), 'Aspects of Scientific Explanation', in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hieronymi, P. (2006) 'Controlling Attitudes,' *Pacific Philosophical Quarterly* 87, no. 1 (2006): 45–74.

--- (2008). 'Responsibility for believing,' *Synthese*, *161*, pp.357-373.

--- (2009). 'Believing at will,' *Canadian Journal of Philosophy*, *39*(sup1), pp.149-187.

--- (2005). 'The Wrong Kind of Reason,' *The Journal of Philosophy* 102, no. 9 (2005): 1–21.

--- (2014). 'Reflection and responsibility,' *Philosophy & Public Affairs*, *42*(1), pp.3-41.

Hills, A. (2016), 'Understanding Why', *Noûs*, 50(4), pp. 661–688.

Hu, X. (2017), 'Why do True Beliefs Differ in Epistemic Value?', *Ratio*, 30(3).

Hyman, J. (2015), *Action, Knowledge, and Will*. New York: Oxford University Press.

James, W. (1896), 'The Will to Believe', *The New World*, 5, pp. 327–347.

Kane, R. (1996), *The Significance of Free Will*. New York: Oxford University Press.

Kavka, G.S. (1983), 'The toxin puzzle', *Analysis*, 43(1), pp. 33–36.

Kornblith, H. (2001), 'Epistemic obligation and the possibility of internalism', in A. Fairweather and L.T. Zagzebski (eds), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. Oxford University Press, pp. 231–248.

Korsgaard, C.M. (1983), 'Two Distinctions in Goodness', *The Philosophical Review*, 92(2), pp. 169–195.

Kripke, S.A. (1980), *Naming and Necessity*. Cambridge: Harvard University Press.

Kvanvig, J.L. (2003), *The value of knowledge and the pursuit of understanding*. Cambridge: Cambridge University Press (Cambridge studies in philosophy).

Latourelle, R. (1968), *Theology of Revelation*. Cork: Mercier Press.

Maimonides, M. (1974), *The Guide for the Perplexed*. 4th edn. Translated by M. Friedländer. New York: E. P. Dutton & Company.

Marušić, B. (2011). 'The ethics of belief,' *Philosophy Compass*, 6(1), 33–43.

McCann, H. (1998), *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca, New York: Cornell University Press.

Mele, A.R. (2003), *Motivation and Agency*. Oxford: Oxford University Press.

--- and Robb, D. (1998), 'Rescuing Frankfurt-Style Cases', *The Philosophical Review*, 107(1), pp. 97–112.

Nottelmann, N. and De Bruin, B., (2021), 'Impermissible Self-Rationalizing Pessimism: In Defence of a Pragmatic Ethics of Belief,' *Erkenntnis*, 86, pp. 257-274.

Pace, M. (2011), 'The Epistemic Value of Moral Considerations: Justification, Moral Encroachment, and James' "Will To Believe"', *Noûs*, 45(2), pp. 239–268.

Peels, R. (2014a), 'Against Doxastic Compatibilism', *Philosophy and Phenomenological Research*, 89(3), pp. 679–702.

--- (2014b), 'Believing at Will is Possible', *Australasian Journal of Philosophy*, 93(3), p. 524.

Perrine, T. (2018), 'Basic Final Value and Zimmerman's The Nature of Intrinsic Value', *Ethical Theory and Moral Practice*, 21(4), pp. 979–996.

--- (2021), 'A puzzle about epistemic value and steps towards a solution', *Synthese*, 199, pp. 12103–12119.

Plantinga, A. and Wolterstorff, N. (eds) (1984), *Faith and Rationality: Reason and Belief in God*. Notre Dame: University of Notre Dame Press.

Plato (2016), *The Republic of Plato*. 2016th edn. Translated by A. Bloom. New York: Basic Books.

Pritchard, D. (2007), 'Recent work on epistemic value', *American Philosophical Quarterly*, 44(2).

--- (2014) 'Truth as the Fundamental Epistemic Good.' In J. Matheson and R. Vitz (eds). *The Ethics of Belief*. Oxford University Press, pp. 112–129

--- (2016a), 'Epistemic Axiology', in M. Grajner and P. Schmechtig (eds), *Epistemic Reasons, Norms and Goals*. De Gruyter, pp. 407–422.

--- (2016b), 'Seeing it for Oneself: Perceptual Knowledge, Understanding, and Intellectual Autonomy', *Episteme*, 13(1), pp. 29–42.

--- (2021), 'Intellectual virtues and the epistemic value of truth', *Synthese*, 198(6).

Pruss, A.R. (2019), 'The Argument from the Denial of Scepticism'. *God, Human Nature and Mathematics*, Wilde Lectures in Natural Theology, Oriel College, Oxford, 16 May.

R v. Hudson and Taylor, 2 All ER 244 (1971).

Rawls, J. (1971), *A Theory of Justice*. Cambridge: Harvard University Press.

Reisner, A. (2018), 'Pragmatic reasons for belief', In Star, D. (Ed.), *The Oxford Handbook of Reasons and Normativity* (Vol. 138/1, pp. 17–27). Oxford: Oxford University Press.

Shabo, S., (2015), 'More trouble with tracing,' *Erkenntnis*, 80, pp.987-1011.

Sharon R. (2010) ‘Doxastic Voluntarism’, in Jonathan Dancy, Ernest Sosa and Matthias Steup (eds.), A Companion to Epistemology. Malden, MA: Wiley-Blackwell.

Shafer-Landau, R. (2003), *Moral Realism: A Defence*. New York: Oxford University Press.

Shah, N. and Velleman, J.D. (2005), ‘Doxastic Deliberation’, *The Philosophical Review*, 114(4), pp. 497–534.

Shepherd, J. (2015), ‘Deciding as Intentional Action: Control over Decisions’, *Australasian Journal of Philosophy*, 93(2), pp. 335–351.

Setiya, K. (2008), ‘Believing at will,’ *Midwest Studies in Philosophy: Truth and its Deformities*, *32*, pp. 36-52.

Sider, T. (2011), *Writing the Book of the World*. Oxford: Oxford University Press.

Siebert, M.K. (2015), ‘Aquinas on Believing God’, *Proceedings of the American Catholic Philosophical Association*, 89, pp. 97–107.

--- (2016), ‘Aquinas on Testimonial Justification: Faith and Opinion’, *The Review of Metaphysics*, 69(3), pp. 555–582.

Smart, J.J.C. (1961), ‘Free Will, Praise and Blame’, *Mind*, LXX(279), pp. 291–306.

Smart, J.A. (2021), ‘Disbelief is a distinct doxastic attitude’, *Synthese*, 198(12), pp.11797-11813.

Sorensen, R. (2011), ‘Interestingly Dull Numbers’, *Philosophy and Phenomenological Research*, 82(3), pp. 655–673.

Sosa, E. (1991), *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.

--- (2003), ‘The Place of Truth in Epistemology’, in M. DePaul and L. Zagzebski (eds), *Intellectual Virtue: Perspectives from Ethics and Epistemology*. New York: Oxford University Press.

Timpe, K. (2011), 'Tracing and the epistemic condition on moral responsibility.' *Modern Schoolman-a Quarterly Journal of Philosophy*, *88*(1), p.5.

Treanor, N. (2014), 'Trivial Truths and the Aim of Inquiry', *Philosophy and Phenomenological Research*, 89(3), pp. 552–559.

Turri, J. (2012), 'A Puzzle About Withholding', *The Philosophical Quarterly*, 62(247), pp. 355–364.

Vargas, M. (2008), 'Moral Influence, Moral Responsibility', in N. Trakakis and D. Cohen (eds), *Essays on Free Will and Moral Responsibility*. Cambridge: Cambridge Scholars Publishing, pp. 90–122.

--- (2005), 'The Trouble with Tracing', *Midwest Studies In Philosophy,* Vol. 29 Issue 1, p269-291.

Velleman, D. (1989), *Practical Reflection*. Princeton, New Jersey: Princeton University Press.

Vlastos, G. (1965), 'Degrees of Reality in Plato', in R. Bambrough (ed.), *New Essays on Plato and Aristotle*. London: Routledge, pp. 1–20.

Weinberg, J.M., Nichols, S. and Stich, S. (2001), 'Normativity and Epistemic Intuitions', *Philosophical Topics*, 29(1/2), pp. 429–460.

Wertheimer, A. (1996), 'Consent and Sexual Relations', *Legal Theory*, 2(2), pp. 89–112.

Whitcomb, D. (2007), *An epistemic value theory*. Rutgers University.

White, N.P. (1992), 'Plato's metaphysical epistemology', in R. Kraut (ed.), *The Cambridge Companion to Plato*. Cambridge: Cambridge University Press, pp. 277–310.

White, R. (2005), 'Epistemic permissiveness', *Philosophical Perspectives*, 19(1), pp. 445–459.

Widerker, D. (1995), 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities', *The Philosophical Review*, 104(2), pp. 247–261.

Wilkenfeld, D.A. (2016), ‘Understanding Without Believing’, in S.R. Grimm, C. Baumberger, and S. Ammon (eds), *Explaining Understanding*. New York: Routledge, pp. 318–344.

Williamson, T. (2002), *Knowledge and Its Limits*. New York: Oxford University Press.

Wittgenstein, L. (2009), *Philosophical Investigations*. 4th edn. Edited by P.M.S. Hacker and J. Schulte. Chichester: Wiley-Blackwell.

Woodward, J.F. (2003), *Making things happen: a theory of causal explanation*. New York: Oxford University Press (Oxford studies in philosophy of science).

Wrenn, C.B. (2017), ‘Truth is not (Very) Intrinsically Valuable’, *Pacific Philosophical Quarterly*, 98(1), pp. 108–128.