

# ON EPISTEMIC AGENCY

A Dissertation Presented

by

KRISTOFFER AHLSTROM

Submitted to the Graduate School of the  
University of Massachusetts Amherst in partial fulfillment  
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2010

Philosophy

© Copyright by Kristoffer Ahlstrom 2010

All Rights Reserved

# ON EPISTEMIC AGENCY

A Dissertation Presented

by

KRISTOFFER AHLSTROM

Approved as to style and content by:

---

Hilary Kornblith, Chair

---

Jonathan Vogel, Member

---

Christopher Meacham, Member

---

Adrian Staub, Member

---

Phillip Bricker, Department Head  
Philosophy

*To Radha*

## ACKNOWLEDGMENTS

Thanks to Richard Feldman, Mikkel Gerken, Alvin Goldman, Stephen Grimm, Klemens Kappel, Christoph Kelp, Michael Lynch, Anne Meylan, Erik Olsson, Nikolaj Jang Lee Linding Pedersen, Casey Perin, Duncan Pritchard, James Simmons, Robert Talisse, and J. D. Trout for helpful feedback on different parts of the present material. Thanks also to Christopher Meacham, Adrian Staub, and Jonathan Vogel for valuable comments and for agreeing to serve on my committee. A special thanks to Hilary Kornblith. I could not have asked for a better philosophical mentor. Last but in no way least, thanks to Radha— for everything.

# ABSTRACT

ON EPISTEMIC AGENCY

SEPTEMBER 2010

KRISTOFFER AHLSTROM, B.A., UNIVERSITY OF GOTHENBURG

Ph. D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hilary Kornblith

Every time we act in an effort to attain our epistemic goals, we express our *epistemic agency*. The present study argues that a proper understanding of the actions and goals relevant to expressions of such agency can be used to make ameliorative recommendations about how the ways in which we *actually* express our agency can be brought in line with how we *should* express our agency. More specifically, it is argued that the actions relevant to such expressions should be identified with the variety of actions characteristic of inquiry; that contrary to what has been maintained by recent pluralists about epistemic value, the only goal relevant to inquiry is that of forming true belief; and that our dual tendency for bias and overconfidence gives us reason to implement epistemically paternalistic practices that constrain our freedom to exercise agency in substantial ways. For example, we are often better off by gathering only a very limited amount of information, having our selection of methods be greatly restricted, and spending our time less on reflecting than on simply reading off the output of a simple algorithm. In other words, when it comes to our freedom to express epistemic agency, more is not always better. In fact, less is often so much more.

# TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	V
ABSTRACT .....	VI
CHAPTER	
INTRODUCTION .....	1
1. WHAT IS EPISTEMIC AGENCY?.....	9
1.1. Agency and Amelioration .....	9
1.2. Doxastic Agency .....	13
1.3. Agency and Rational Authority .....	17
1.4. On the Epistemic Merits of Reflection .....	23
1.5. Is Epistemic Agency a Myth? .....	28
1.6. Conclusion .....	32
2. I CAN'T MAKE YOU LOVE TRUTH (AND YOU CAN'T MAKE ME LOVE EVIDENCE).....	35
2.1. Goals and Values .....	35
2.2. Feldman on Epistemic Value .....	39
2.3. Is Feldman Mistaken About Value? .....	43
2.4. True Belief and Human Flourishing .....	46
2.5. Epistemic Value and the Ethics of Belief .....	52
2.6. Conclusion .....	58
3. IN DEFENSE OF EPISTEMIC VALUE MONISM .....	60
3.1. Varieties of Value .....	60
3.2. On the (Alleged) Surplus Value of Knowledge .....	67
3.3. Meno and the Monist .....	75
3.4. Justification and Non-instrumental Value .....	86
3.5. The Trials and Achievements of Understanding .....	96
3.5.1. Objectual Understanding .....	96
3.5.2. Understanding-Why .....	102
3.6. Conclusion .....	107
4. A FRAMEWORK FOR EVALUATION.....	108
4.1. Defining Epistemic Betterness.....	109
4.1.1. Questions and Informative Answers .....	111
4.1.2. Problems of Individuation .....	115
4.1.3. Shared Agendas .....	119
4.1.4. On Incomparability.....	120

4.2. On the Specialized Role of Epistemic Evaluation .....	123
4.2.1. Degrees of Interest.....	125
4.2.2. Banal and Harmful Pursuits.....	128
4.3. Conclusion .....	133
5. EPISTEMIC AGENCY AND PATERNALISM: WHEN LESS IS MORE.....	134
5.1. Reasons for Mandating Compliance.....	134
5.2. What is Epistemic Paternalism?.....	140
5.3. Two Cases for Expansion .....	147
5.3.1. Prediction Models in Diagnosis and Prognosis .....	148
5.3.2. Prediction Markets as an Alternative to Social Deliberation .....	152
5.4. Countervailing Reasons .....	158
5.4.1. Reasons for Not Relying on Prediction Models .....	159
5.4.2. Reasons for Not Relying on Prediction Markets.....	162
5.5. Conclusion .....	167
BIBLIOGRAPHY .....	168



## INTRODUCTION

When considering the last fifty years of epistemology in the analytical tradition—filled with far-fetched thought experiments involving evil demons, reliable clairvoyants, and disguised mules—it is easy to forget that there is an important sense in which epistemology is an utterly practical discipline. Indeed, according to a long-standing tradition, one of epistemology’s main missions is to provide hands-on advice, aiding the epistemic inquirer in her pursuits. Understood thus, epistemology is not only *normative*, in the sense that it concerns itself with specifically epistemic goods, but also *ameliorative*, in that it attempts to say something constructive about how our odds of actually attaining the relevant goods may be increased.

Ameliorative epistemology is not a new thing, nor is it “a departure from the central issues taken up in epistemology,” as has been suggested by Richard Feldman (1999, p. 172). To the contrary, it was practiced by two of the most central historical figures in Western epistemology, i.e., John Locke and Rene Descartes. The ameliorative ambition of the latter is most obvious in his *Rules for the Direction of the Mind* and *Discourse on the Method*—two works explicitly in the business of epistemic prescription. Similarly, one of Locke’s main concerns was the reformation of people’s intellectual lives from one based on tradition to one based on reason. Part of the motivation for such a reformation was, as Locke (1706/1996) writes in the posthumously published *Of the Conduct of the Understanding*, simply that “there are a great many natural defects in the understanding capable of amendment” (§2). In addition, bringing about the relevant kind

of amendment would, as Locke saw it, serve as a protection against the social conflicts resulting from the breakup of medieval Christendom's intellectual consensus.<sup>1</sup>

As the last fifty years of experimental psychology makes clear, however, amelioration is a worthwhile endeavor even in the absence of an imminent threat of large-scale, social turmoil. The often banal ways in which our dual tendencies for bias and overconfidence constantly get in the way of accuracy and informed action—unfortunately, often with far from banal consequences—is a sufficient reason for trying to ameliorate our epistemic conduct and practices. On a more positive note, the very same body of empirical research provides us with something that our epistemological ancestors lacked, namely robust data on the systematic and often predictable ways in which we go astray, as well as actionable evidence regarding what kinds of ameliorative recommendations are likely to be effective.

This is *not* to suggest that the epistemologist should be expected to do the job of the empirical psychologist.<sup>2</sup> Instead, it is to seek a middle ground between those who would suggest that epistemology just *is* empirical psychology (as some uncharitable readings of Quine would have him say) and those who wish to engage in amelioration without at all engaging with the relevant empirical literature. As for the latter, consider Robert Roberts and Jay Wood's recent defense of what they call *regulative epistemology*—a kind of epistemology with explicit, ameliorative ambitions:

[...] to say that our virtue epistemology is regulative is not to deny that it's analytic. In fact, what we call analysis is our chief expedient of regulation. By the

---

<sup>1</sup> See Wolterstorff (1996).

<sup>2</sup> That is, unless the philosophers in question are trained to do so. In this context, consider the interesting and in many ways commendable cross-scientific research associated with so-called experimental philosophy.

conceptual work that is distinctive of philosophical discourse, we propose to facilitate the improvement of intellectual character. If conceptual analysis is done right, it clarifies the character of the intellectual life in a way that can actually help people live that life. Conceptual clarification is an important part of education, and the improvement of intellectual character is a kind of education. Conceptual clarification is not the whole of education [but] it is at least something, and it is what the philosopher is well suited to contribute (Roberts and Wood 2007, pp. 27-28).

Roberts and Wood's commitment to epistemic improvement is commendable. But the manner in which they restrict their contribution to that of conceptual analysis (as evidenced by their book at large, not solely by the above passage) seems unnecessarily limiting. Granted, philosophers receive ample training in conceptual clarification, which is, clearly, something that comes in handy in all forms of inquiry. And it would probably be a bad idea to assume that the skill-set of the typical philosopher naturally translates into an aptitude for experimental work. But it is one thing to suggest that philosophers are not well-suited for contributing empirical work of their own, and quite another to fail to consult the vast amount of empirical work that is of high relevance to the multitude of empirical bets—perhaps most pertinently, regarding what means of regulation will be conducive to the relevant ends, given the surprising ways in which cognition sometimes works—made in the course of any attempt to regulate epistemic behavior or shape epistemic institutions. In fact, to be in the business of telling people how to think without consulting the relevant empirical work borders on the irresponsible.

In contrast to the project of Roberts and Wood, the present study places itself firmly within the ameliorative tradition of Locke and Descartes, as it has been developed more recently within the naturalist movement, combining a commitment to amelioration with the acknowledgement that such amelioration needs to be informed by the

psychology of cognition in order to be effective.<sup>3</sup> While much has been said about this naturalistic movement,<sup>4</sup> the motivation for the particular investigation to be undertaken below is that one notion in particular has received little to no attention in connection with the ameliorative project, i.e., the notion of *epistemic agency*. This omission is unfortunate since the notion of epistemic agency has the potential of playing a central role in ameliorative epistemology—or so it will be argued.

But first, what is epistemic agency? As will be suggested in chapter 1, epistemic agency is the kind of agency we express through actions performed in an effort to attain our epistemic goals. This characterization begs for clarification on at least two points. (Further clarifications will be postponed for now.) First, what are the *actions* by way of which we express such agency? Second, what are our *epistemic goals*, or goal, if it turns out that there is only one? Chapter 1 approaches the first question by considering three candidate accounts of the relevant actions. The first account identifies epistemic agency with *doxastic* agency, or decisions about what to believe. The second account identifies epistemic agency with the kind of *reflective* agency or authority that some have maintained that we have over our reflectively formed beliefs. The third account also identifies epistemic agency with reflective agency, but motivates it with reference to the epistemic benefits of reflecting.

All three accounts are rejected, the first one because it is committed to doxastic voluntarism, the second one because it fails to delimit a sufficiently robust notion of freedom to be relevant to amelioration, and the third one because it relies on too optimistic a view of the epistemic merits of reflection. Contrary to other recent critiques

---

<sup>3</sup> See, e.g., Bishop and Trout (2005), Kitcher (1992) and Goldman (1978).

<sup>4</sup> See Kornblith (1994).

of the relevant accounts, however, it will not be concluded that the appeal to epistemic agency, thereby, is “nothing more than a bit of mythology.”<sup>5</sup> The notion of epistemic agency is a useful one. More specifically, if we take epistemic agency to be expressed through actions performed in an effort to attain our epistemic goals, a proper understanding of the domain of epistemic agency serves to delimit exactly the kind of activities with which ameliorative recommendations are to be concerned.

In fact, understood thus, the question of what actions are relevant to the expression of epistemic agency is almost trivial. After all, to perform actions for the purpose of forming true belief in a manner that furthers our epistemic goals is to engage in *inquiry*. In other words, epistemic agency is constituted by all the things we do when conducting inquiry. We gather information, mull over our data, choose among different methods of investigation, and so on, and in so far as we are doing all of this in an attempt to attain our epistemic goals, we are expressing our epistemic agency. At the same time, the very triviality of such an account of the actions relevant to epistemic agency suggests that the interesting questions regarding epistemic agency lie elsewhere. One such question is the ameliorative question, namely: How can the ways in which we *actually* express our agency be brought in line with how we *should* express our agency, given our epistemic goals? However, this question begs a further one, posed already in the above, namely: What are our epistemic goals?

This gets back to the observation that epistemology is normative. To talk about epistemic goals is to talk about epistemic value, in that the goals of inquiry determine what activities, states, processes, practices, and so on, are epistemically valuable.

---

<sup>5</sup> Kornblith (forthcoming, p. 118; all page references to this work refer to the manuscript).

Chapters 2 and 3 develop and defend a particular account of such value that has come under fire in recent discussions in epistemic axiology, namely that true belief is the only goal of inquiry and, consequently, also unique in being of non-instrumental epistemic value. More specifically, the idea that believing truly is valuable thus is defended on two fronts, both from those denying that true belief is *a* non-instrumental epistemic value, and from those who deny that true belief is the *only* non-instrumental epistemic value.

Chapter 2 takes on the former challenge by considering the case for a kind of evidentialism that takes believing upon sufficient evidence to be the only goal of inquiry and, consequently, also the only thing of non-instrumental epistemic value. For reasons that will be discussed below, the strategy will not be to reject such a take on value, but merely to protect the idea that truth is of non-instrumental epistemic value from the evidentialist's contention that it is not. More specifically, it will be argued that the prospects for providing someone who does not share one's love for the relevant non-instrumental epistemic goods with reasons for starting to do so seem dim. However, the point applies equally to the evidentialist as to those of us who value true belief non-instrumentally from an epistemic point of view. Moreover, given that failure to convince someone to value something is by no means a sufficient reason for ceasing to value that thing, this dialectical fact renders neither axiological position illegitimate. In light of this, subsequent chapters assume the position taken by most epistemologists to the effect that true belief is, indeed, a non-instrumental epistemic good.

Chapter 3 takes up the second challenge, i.e., that of showing that true belief is not just one non-instrumental epistemic value among many. Here, the strategy will be less reconciliatory. Recently, epistemic value monism—i.e., the idea that believing truly

is unique in possessing non-instrumental epistemic value—has come under attack by philosophers arguing that we cannot account fully for the domain of epistemic value in monistic terms. However, it will be argued that the relevant critiques fail to establish any such thing. For one thing, there is a presumption of monism due to considerations about axiological parsimony. Granted, such a presumption would be defeated by positive evidence to the effect that the relevant kind of monism makes us unable to fully account for the domain of epistemic value. But a proper examination and subsequent rebuttal of the most promising counterexamples to monism casts serious doubt upon the claim that there is any such evidence. Consequently, epistemic value monism still stands, and true belief is not only a but the *only* good of non-instrumental, epistemic value.

It does not follow from true belief being the only good of non-instrumental epistemic value that *all* true beliefs are valuable thus. More specifically, chapter 4 argues that the epistemic value of a practice is a function of the extent to which it yields true beliefs that answer questions posed by relevant sets of inquirers. As such, the account tells us something about what methods or practices inquirers *should* utilize, in so far as they want answers to their questions. The remainder of the chapter is primarily concerned with tracing the limits of this ‘should,’ and suggests that there are very real limits to the extent to which we can make *epistemological* judgments about what questions inquirers should pose (as opposed to about what practices to opt for, *given* a set of questions). However, it is also argued that, rather than indicating a flaw in the account, this simply suggests that it is able to account for the specialized nature of epistemic evaluation.

Armed with a framework for epistemic evaluation, chapter 5 returns to the ameliorative question by illustrating the manner in which a proper understanding of the

actions and goal relevant to epistemic agency can be used to make recommendations about how the ways in which we actually express our agency may be brought in line with how we should express our agency. More specifically, a case is made for the claim that our dual tendency for bias and overconfidence gives us *pro tanto* reason for mandating compliance with methods that have been shown to increase our reliability. Moreover, it is argued that mandating compliance thus would be *epistemically paternalistic*, and that we have reason—indeed, not just *pro tanto* but all-things-considered reason—to practice such paternalism on a wider scale than we are already doing.

To sum up, when properly understood, epistemic agency has nothing to do with decisions about what to believe, and far less to do with reflection than some have suggested. Instead, epistemic agency is constituted by all the things we do when conducting inquiry. As such, epistemic agency is neither a myth, nor something the expression of which should be considered epistemically valuable in its own right. After all, given the plausibility of epistemic value monism and the prevalence of bias and overconfidence, it turns out that we are often better off from the point of view of attaining our epistemic goal by having our freedom to exercise agency be constrained by paternalistic practices. In other words, when it comes to our freedom to express agency, more is not always better. In fact, less is often so much more.



## CHAPTER 1

### WHAT IS EPISTEMIC AGENCY?

#### 1.1. Agency and Amelioration

What is it to express epistemic agency? It is, clearly, to *do* something. But, surely, that is not sufficient. Someone falling down a set of stairs is doing something, namely falling down a set of stairs, but not, thereby, expressing any agency, let alone any *epistemic* agency. So, perhaps to express epistemic agency is to engage in a kind of *goal-directed* action. But the relevant kind of action cannot be directed toward just *any* goal. A stunt man deliberately and skillfully falling down a set of stairs on a movie set is not only doing something but also performing an action directed at the goal of making it seem as if she is falling down the stairs by accident, while doing so deliberately and (if skillfully) with no significant risk of harm. So, what is the particular goal—or goals, if there turns out to be several—toward which the kind of actions characteristic of *epistemic* agency needs to be directed? Subsequent chapters will make a case for a particular answer to this question, namely true belief, and use it to provide the foundation for an account of epistemic value. For the remainder of this chapter, however, this answer will simply be taken for granted.

Even if we assume that epistemic agency is expressed by way of actions directed at the goal of forming true belief, several questions remain. First, what does it mean to be “directed” at the goal of forming true belief? Does it imply that the agent in question *consciously* entertains that goal whenever performing whatever actions are associated

with expressions of epistemic agency? That seems too strong. Many paradigmatic instances of what we would intuitively want to categorize as expressions of epistemic agency, such as the everyday inquiry of a scientific researcher, involve subjects that do not consciously keep that goal in mind at all times. Surely, they *could* bring it to mind, if properly prompted, and if we had to put any constraint on agency it would be that the agent in question at least has the ability to entertain (and endorse) the relevant goal consciously, although there is no need for her to actually do so in order to qualify as an epistemic agent. Such a constraint would rule out what seem less controversial instances of non-agency, such as that of the infant crawling across the floor in proto-inquiry, or an ant foraging. This is not to suggest that epistemic agency is a prerogative of adult humans. Some adults might lack the conceptual sophistication required to think of themselves as engaging in a project directed at forming true belief, and some non-human animals may be sophisticated enough to qualify as agents.

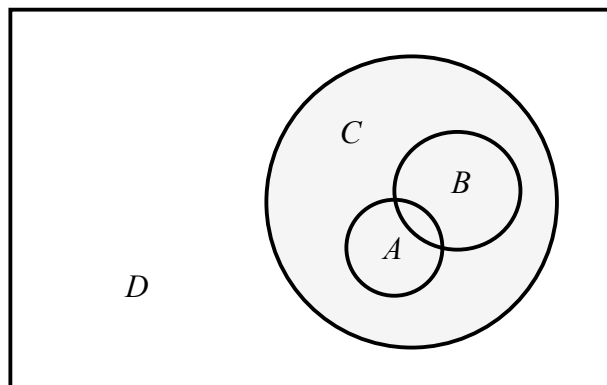
Second, do the actions directed at the goal of forming true belief need to be *successful*, i.e., actually terminate in the attainment of true belief, in order to qualify as expressions of epistemic agency? In one sense, clearly not. A researcher has not failed to express epistemic agency just because she gets something wrong. That said, our intuitions get less clear when we consider the other limiting case, i.e., someone sincerely pursuing truth but doing a terrible job at it. Consider, for example, an astrologist who sincerely believes that her study of the movements and relative positions of celestial bodies provides her with predictive data about human behavior. Could she (*qua* astrologist) still qualify as an epistemic agent? In so far as we are inclined to say “no,” perhaps on account of how her ideas about what is truly predictive of what are so far off the mark,

we might want to put the following success constraint on epistemic agency: in order to qualify as an epistemic agent, one's methods of inquiry must be generally reliable with respect to the relevant subject matters. Let us refer to a notion of agency invoking this constraint as a *strong* notion of agency.

I hesitate to embrace such a strong notion because it places those in greatest need of intellectual advice outside the scope of agency. Moreover, rejecting above success constraint and, in effect, opting for what we may call a *weak* notion of agency, does not serve to disqualify considerations about success from questions that may be posed in terms of epistemic agency. On such a weak notion, aforementioned astrologist could qualify as an epistemic agent; she would just be a terrible epistemic agent. Given that diagnosis, we could admit her into the domain of epistemic agency, and then determine how she could start doing things differently, in order to make more successful predictions about human behavior. After all, if you are interested in providing useful intellectual advice, what you are interested in doing are two things. First, you want to determine what people engaging in actions directed (successfully or not) at the goal of forming true belief are *actually* doing. Second, to the extent that they are unsuccessful in their pursuits, you want to say something constructive about what the agents in question *should* be doing differently in order to actually attain that goal.

This provides the first motivation for an inquiry into the notion of epistemic agency: Contrary to what its absence in previous epistemological treatments might lead us to believe, the notion of epistemic agency has an important role to play in ameliorative epistemology in that a proper understanding of the domain of epistemic agency serves to delimit exactly the kind of activities with which ameliorative recommendations are to be

concerned. One way to illustrate this point in more detail is by thinking of the domain of actions relevant to epistemic agency in terms of a simple Venn diagram, as follows:



*A* encompasses the ways in which we *actually* express our agency. These actions are relevant because, as noted above, our actual epistemic practices provide the point of departure for any ameliorative advice. *C* encompasses the ways in which we *could* express our agency, and  $C-A$  the ways in which we could express our agency differently from what we are presently doing. Furthermore,  $A \cap B$  designates the ways in which we are doing things *right* from the point of view of attaining true belief, and  $B-A$  the ways in which we could be doing things differently and for the *better*. Finally, everything that falls outside the scope of *C* pertains to the ways in which we could *not* express our agency, designated above by *D*. As will become clear in the next section, *D* is relevant for a purely negative reason in that it tells us something about the ways in which we cannot do things differently, and what we, consequently, have to work around rather than with, as far as intellectual advice is concerned.

The second motivation for an inquiry into the notion of epistemic agency is what most likely also provides the reason for the notion's absence from the relevant discussions, namely a series of accounts about what epistemic agency amounts to that are

of no help in so far as we are interested in amelioration. It is the burden of the remainder of this chapter to consider some of the most prominent accounts of such agency, demonstrate the ways in which they fall short as far as the ameliorative project is concerned, and then show how the notion of epistemic agency can be re-established so as to play an important role within that project, in line with what was just outlined.

## 1.2. Doxastic Agency

Above, it was argued that to express epistemic agency is to engage in actions directed (consciously or not) at the goal of forming true belief. What are the relevant *actions*? One way to approach this question is by considering one of the most fundamental objects of epistemic evaluation, i.e., belief. Beliefs can be justified, rational, and constitute knowledge. In addition, our beliefs make up the “map of neighbouring space by which we steer,” as pointed out by Frank P. Ramsey (1931). Given that some ways of steering are more successful than others (epistemically or otherwise), one straightforward way to understand the actions of epistemic agency is in terms of *decisions* about what beliefs to form (or not to form). Let us refer to the relevant kind of agency—if there is, indeed, such agency—as *doxastic* agency.

The main problem with the suggestion that epistemic agency is to be identified with doxastic agency is that it amounts to assuming *doxastic voluntarism*. That is, it amounts to assuming that belief-formation is under our voluntary control and that we, in effect, are able to form beliefs at will. This view seems doubtful, to say the least.

“Believing, we feel, is not something one can in any sense *decide* to do or refrain from, it

seems rather a *condition* in which one finds oneself,” as John Heil writes (1984, p. 56).

William Alston elaborates on a similar point:

Can you, at this moment, start to believe that the Roman Empire is still in control of Western Europe, just by deciding to do so? If you find it incredible that you should be sufficiently motivated to even try believing this, suppose that someone offered you \$500 million to believe it, and that you are much more interested in the money than believing the truth. Could you do what it takes to get at the reward? Remember that we are speaking of believing *at will*. No doubt, there are things you could do that would increase the probability of your believing this [but] can you switch propositional attitudes toward that proposition just by deciding to do so? It seems clear to me that I have no such powers. Volitions, decision, or choosings don't hook up with propositional attitude inaugurations, just as they don't hook up with the secretion of gastric juices or with metabolism (Alston 2005, p. 63).

At this point, the voluntarist may object that the aforementioned line of reasoning presupposes a far too strong reading of “voluntary control.” More specifically, it might be objected that it is being assumed that the relevant kind of voluntarism requires *direct* control over our belief-forming processes, where having direct voluntary control implies being able to bring about the object of control by a mere act of will, intention or volition.<sup>1</sup> This is the kind of control we might have over mental imagery—i.e., our ability to bring about mental pictures—but that we do not seem to have over our belief-forming processes. However, the voluntarist may insist that there are other forms of control, most pertinently various forms of *indirect voluntary control*, located on a continuum of decreasingly direct control. For example, certain things may be brought about not by a direct act of will, but by way of a chain of events, making up a single, uninterrupted act. This is the sense in which opening my door or adjusting the temperature in my office is

---

<sup>1</sup> The distinction between direct and indirect voluntary control is borrowed from Alston (2005), as are the further distinctions between different kinds of control discussed below.

within my voluntary control; while I may not be able to open the door or alter the temperature merely by an act of will, the opening of the door and altering of the temperature is nevertheless within my (indirect) voluntary control, by virtue of it being possible for me to bring about the necessary chain of events that is me getting up, walking across the room, and opening the door or turning the knob on the thermostat.

Even further out on the continuum of indirect voluntary control are the kinds of acts that do not constitute single, uninterrupted acts, like walking over and opening a door or turning a knob, but rather a series of inter-related actions spread out over an extended period of time. This is the sense in which looking for more evidence, taking steps to engage in more thorough consideration and weighing of evidence, and deliberating over how to direct my inquiry in a more strategic manner is within my *long-range (indirect) voluntary control*. This is also the kind of control Roderick Chisholm seems to be getting at in the following passage:

If self-control is what is essential to activity, some of our beliefs, our believings, would seem to be acts. When a man deliberates and comes finally to a conclusion, his decision is as much within his control as is any other deed we attribute to him. If his conclusion was unreasonable, a conclusion he should not have accepted, we may plead with him: "But you needn't have supposed that so-and-so was true. Why didn't you take account of these other facts?" We assume that his decision is one he could have avoided and that, had he only chosen to do so, he could have made a more reasonable inference. Or, if his conclusion is not the result of a deliberate inference, we may say, "But if you had only stopped to think", implying that, had he chosen, he could have stopped to think. We suppose, as we do whenever we apply our ethical or moral predicates, that there was something else the agent could have done instead (Chisholm 1968, p. 224).

As pointed out by Alston, however, any attempt to frame doxastic voluntarism in terms of indirect voluntary control (long-range or otherwise) needs to ignore a crucial distinction,

namely the distinction between doing *C* to voluntarily bring about an *E*, and doing *C* to voluntarily bring about a *definite E*. Alston elaborates:

In order that the phenomenon of looking for more evidence would show that we have voluntary control over propositional attitudes, it would have to be the case that the search for evidence was undertaken with the intention of taking up a *certain* attitude toward a *specific* proposition. For only in that case would it have any tendency to show that we have exercised voluntary control over *what* propositional attitude we come to have (Alston 2005, p. 70).

In other words, the fact that we do have voluntary control over many actions that might give rise to beliefs does not show that we have voluntary control over *what* beliefs we form as a result of those actions. The same goes for various forms of *indirect voluntary influence*—a kind of influence even further out on the continuum of decreasing directness—where we may take steps that either bring to bear on candidates for belief or on our general belief-forming habits and tendencies. That we may have such influence over our beliefs does *not* show that *doxastic* voluntarism is true, i.e., that we may voluntarily choose what *specific* propositions to believe, anymore than our ability to open our eyes and look out a window on a sunny day shows that we may choose whether or not to believe that it is sunny.

The implications of this are far from trivial as for how we are to understand epistemic agency, since the failure of doxastic voluntarism suggests that the domain of belief-formation not only does not but *cannot* coincide with the domain of epistemic agency. Indeed, above considerations suggest that the two domains are completely distinct, since one pertains to something that we *do* (i.e., agency), and the other to something that simply *happens* to us (i.e., belief-formation). More than that, if the two domains are incorrectly identified, the prescriptive import of epistemic agency, as it



relates to epistemic amelioration, will be reduced to a set of unhelpful intellectual recommendations in terms of belief and its suspension—recommendations that, to borrow Alston’s analogy, are no more helpful than dietary recommendations framed exclusively in terms of the mechanisms of our digestive apparatus.

### 1.3. Agency and Rational Authority

At this point, it might be objected that the previous section invoked a too narrow notion of our doxastic activities, as one involving mere responses to external stimuli. After all, as human beings, we not only form first-order beliefs, such as perceptual belief; we also *reflect* on our first-order beliefs, as a consequence of which we form certain reflective, second-order beliefs, or beliefs about our beliefs.<sup>2</sup> More than this, unlike belief, reflection is not simply a condition that we find ourselves in; it is something that we do. For one thing, we may choose when and on what to reflect—a fact that will be considered in the next section. For another, the phenomenology of reflection surely suggests that we have *some* say when it comes to what beliefs we end up with as a result of reflecting.

The latter idea is central to Richard Moran (2001).<sup>3</sup> According to Moran, reflective belief-formation, or what he calls *deliberative* belief-formation, is in an important sense *up to us*. However, he explicitly rejects the idea that he is, thereby, committing himself to any form of doxastic voluntarism: “The agency a person exercises

---

<sup>2</sup> ‘Reflection’ is sometimes used in a manner that does not (necessarily) involve any second-order beliefs, as when we talk about reflecting on our data. That is not how the notion of reflection will be used here.

<sup>3</sup> Similar themes to those found in Moran can be found also in Korsgaard (1996), on which parts of Moran’s investigation rest. However, given that Moran is more explicitly concerned with belief-formation than Korsgaard is, I will focus on the former presently.

with respect to his belief and other attitudes,” he writes, “is obviously not like that of overt basic actions like reaching for a glass” (p. 114). Still, we have a very real *authority* over what beliefs we form, and this authority is what makes it possible for the agent to be responsible for her beliefs:

Beliefs and other attitudes [...] are stances of the person to which the demand for justification is internal. And the demand for justification internal to the attitudes involves a sense of agency and authority that is fundamentally different from the various forms of direction or control one may be able to exercise over some mind or other (Moran 2001, p. 114).

It should be noted that Moran’s concern with a *sense* of agency and authority, albeit, clearly, inspired by the phenomenology of reflection, should not be interpreted *merely* as a point about phenomenology. Granted, at times, Moran talks as if the authority in question is simply a matter of the agent *assuming* responsibility for her actions (e.g., on p. 131)—a way of framing the issue that is compatible with such assumptions being metaphysically mistaken. However, the great majority of passages makes clear that Moran is committing himself to a stronger claim, in taking it that reflective belief-formation not only *seems* to be but really *is* up to us. Indeed, Moran’s book is fraught with locutions to the effect that we *make up our minds* (p. 92) by *stepping back* from some content of consciousness (p. 141) in order to *call into question* what we are inclined to believe (p. 143) and *decide* or *commit* ourselves to believing something (p. 145).

So, how are we to interpret these locutions, without either committing Moran to doxastic voluntarism, or taking them to be mere comments on the phenomenology of reflection? Consider the following passage from Moran, succeeding a discussion of Anscombe’s notion of intention:

The stance from which a person speaks with any special authority about his belief or his action is not a stance of causal explanation but the stance of rational agency. In belief as in intentional action, the stance of the rational agent is the stance where reasons that justify are at issue, and hence the stance from which one declares the authority of reason over one's belief and action. Anscombe's question "why?" is asking not for what might best explain the movement that constitutes the agent's action, but instead is asking for the reasons he takes to justify his action, what he is aiming at. It is as an expression of the authority of reason here that he can and must answer the question of his belief or action by reflection on the reasons in favor of this belief or action. To do otherwise would be for him to take the course of his belief or his intentional action to be up to something other than his sense of the best reasons, and if he thinks *that*, then there's no point in his deliberating about what to do. Indeed, there is no point in calling it "deliberation" any more, if he takes it to be an open question whether this activity will determine what he actually does or believes. To engage in deliberation in the first place is to hand over the question of one's belief or intentional action to the authority of reason (Moran 2001, p. 127).

Understood thus, there might, at first glance, seem to be a tension between the *rationality* and the *agency* of rational agency. After all, how do we square the idea of agency or freedom with the restrictive demands of reason and rationality? To understand Moran's answer, it might help to consider Descartes' answer to a question not too different from this one. In his *Meditations on First Philosophy*, Descartes argues that God has bestowed us with free will. At the same time, Descartes takes it that the purest form of freedom is the one exercised when the clarity and distinctness of a truth is so apparent that it prompts more or less automatic assent, and the will is completely overpowered, so to speak, by the psychological force of the intellect.<sup>4</sup> At first, this might seem incompatible with freedom, but Descartes denies that it is:

[...] the will simply consists in our ability to do or not to do something (that is, to affirm or deny, to pursue or avoid); or rather, it consists simply in the fact that when the intellect puts something forward for affirmation or denial or for pursuit

---

<sup>4</sup> See Descartes (1641/1984, p. 48; AT VII 69).

or avoidance, our inclinations are such that we do not feel we are determined by any external force. In order to be free, there is no need for me to be inclined both ways; on the contrary, the more I incline in one direction—either because I clearly understand that reasons of truth and goodness point that way, or because of a divinely produced disposition of my inmost thoughts—the freer is my choice (Descartes 1641/1984, p. 40; AT VII 57-58).

In other words, to be free in the relevant sense is not to be able to opt out of the demands of rationality, but rather to reason in ways that are sensitive to the demands in question. In fact, according to Descartes, freedom takes its purest form when perfectly aligned with those demands. Moran seems to be working with a relevantly similar notion of freedom, as made clear in the following passage, where he discusses the relevant notions of control, activity and responsibility, as they pertain to desires and belief (which, for present purposes, are perfectly analogous, according to Moran):

The person's responsibility [...] is to make his desire answerable to and adjustable in the light of his sense of some good to pursue. It is not a responsibility that reduces to the ability to exert influence over one's desires, and that is why the idiom of "control" is misleading in this context. When the desire is (already) the expression of the person's reasons, there is no *need* for exerting any control over it. As in the case of ordinary theoretical reasoning, which issues in a belief, there is no further thing the person *does* in order to acquire the relevant belief once his reasoning has led him to it. At the beginning of his practical reasoning he was not aiming to produce a particular desire in himself (as he might with respect to another person), but rather holding open his desire to how the balance of reasons falls out (Moran 2001, pp. 118-119).

In other words, the relevant kind of authority, i.e., *rational* authority (p. 124), is one on which we are not so much choosing what to reflectively believe—as in: choosing what particular propositions to believe (or refrain from believing)—as simply leaving it up to our reflective faculties to lead us toward the proper conclusion, under the assumption that they are sensitive to the demands that rationality places on us. That is why Moran's

account of agency is compatible with doxastic involuntarism; the kind of freedom or agency associated with rational authority does not require that we can decide *what* beliefs to form on the basis of reflection.

That said, if this is Moran's take on what it is for the beliefs formed on the basis of reflection to be "up to us," his account of the kind of authority at issue is not relevant to epistemic agency, in so far as the latter notion relates to epistemic amelioration. To see why, return to the first motivation for an inquiry into the domain of epistemic agency discussed above. What we noted was that, in so far as we are interested in providing intellectual advice, we are interested in finding out two things, namely what people engaging in actions directed at the goal of forming true belief are actually doing, as well as what they should be doing differently, in so far as they could be doing better with respect to attaining that goal. Given that the objective of amelioration is to provide intellectual advice, that someone should do something differently presupposes that the target activity is one over which she has effective voluntary control—which was exactly what disqualified doxastic agency as a plausible account of agency. Again, in want of any voluntary control over our beliefs, ameliorative recommendations cannot be formulated in terms of decisions about what to believe, any more than dietary recommendations can be framed exclusively in terms of the mechanisms of our digestive apparatus.

If Moran is right, we do have a kind of authority over our reflective beliefs, in so far as our beliefs are formed under the rational authority of reason. As we have also seen, however, Moran makes it clear that the kind of authority manifested under such circumstances neither requires nor typically involves any voluntary control (let alone any *effective* voluntary control) on part of the deliberating subject. While this

acknowledgement has him steer clear of doxastic voluntarism, it also serves to insulate the relevant reflective processes from ameliorative recommendations. After all, to have authority over what you believe upon reflection is to hand over your belief-formation to the authority of reason (as Moran puts it), and to do so is exactly to *not* be able to believe otherwise—nor, consequently, be able to be effectively *advised* to believe otherwise—unless your reasons require you to believe otherwise.

Although this suggests that there is no agency of the kind relevant to amelioration to be found in what Moran refers to as rational authority, it is important to note that this does not serve to discredit his account, as far as his (non-ameliorative) project is concerned. When encountering evidence that things are not going my way, I may have a strong desire to believe contrary to what the evidence suggests. If I could believe (and not believe) at will upon reflection, resisting the force of the evidence would be an easy feat. That, however, does not change the fact that there is a sense in which I would not be believing as I should, were I to resist the evidence thus. Differently put, as far as epistemology is concerned, our belief-forming tendencies—reflective or not—should not be sensitive to what we *want* to believe, but rather to what we have *reason* to believe. So, perhaps it is not such a bad thing that our belief-forming tendencies are to a certain extent (albeit not completely, as instances of wishful thinking go to show) insulated from our desires and the influence of the will—which, of course, is exactly in line with Moran’s account of rational authority.

#### 1.4. On the Epistemic Merits of Reflection

The previous section suggested that, in so far as there is a sense in which what beliefs we end up with as a result of reflection is up to us, this does not serve to delimit a notion of agency relevant to amelioration. However, above discussions largely ignored a perfectly legitimate domain of agency associated with reflection, identified in passing above when it was noted that we may choose voluntarily when and on what to reflect (although, as we have seen, not what beliefs to form as a result of such choices). To this extent, reflection, clearly, is an action relevant to epistemic agency, since we can engage in it in a manner that is directed at the goal of forming true belief. The questions to be discussed in this section, however, is whether it is the *only* thing that we can do thus, or even something that it generally is a *good* thing to do, given the goal in question. Let us attend to the latter issue first.

Why is it a good thing to reflect? According to Ernest Sosa,

[...] reflection aids agency, control of conduct by the whole person, not just by peripheral modules. When reasons are in conflict, as they so often are, not only in deliberation but in theorizing, and not only in the higher reaches of theoretical science but in the most ordinary reasoning about matters of fact, we need a way holistically to strike a balance, which would seem to import an assessment of the respective weights of pros and cons, all of which evidently is played out through a perspective on one's attitudes and the bearing of those various reasons (Sosa 2004, p. 292).

Notice that Sosa's claim that reflection aids agency depends, crucially, on certain assumptions about the epistemic merits of reflection. After all, it would make little sense to talk about reflection aiding agency, if it turned out that reflection generally did not weigh pros and cons *properly*, nor strike an epistemically *appropriate* balance between reasons. If that turned out to be the case, I highly doubt that Sosa would still be inclined

to consider reflection *aiding* agency, in the relevant sense.<sup>5</sup> In short, in so far as reflection aids agency, it does so in virtue of the epistemic merits of reflection.

This, moreover, is completely in line with what Sosa (1991) says in relation to his account of reflective knowledge. According to Sosa, reflecting on your beliefs, and, thereby, not only demonstrating a “direct response to the fact known but also [an] understanding of its place in a wider whole that includes one’s belief and knowledge of it and how these come about” (p. 240) will, at least in a preponderance of cases, make you epistemically better off than not reflecting on your beliefs.<sup>6</sup> Distinguishing reflective from animal knowledge, i.e., the kind of knowledge resulting from a cognitive faculty that is reliable with respect to a particular field of propositions, Sosa writes:

Since a direct response supplemented by such understanding would in general have a better chance of being right, reflective knowledge is better justified than corresponding animal knowledge (Sosa 1991, p. 240).

If Sosa is right, it would seem reasonable to have reflection play an important role in our account of epistemic agency, since reflection would not only pertain to something we can do but also to something the exercise of which serves the goals of such agency well. And as we have seen, Sosa does, indeed, take reflection to be central to epistemic agency.

---

<sup>5</sup> This is further evidenced by the fact that Sosa (in conversation) denies that that the kind of empirical evidence to be discussed below, suggesting (it will be argued) that reflecting does *not* generally make one better off from an epistemic point of view, gives us any reason to re-evaluate the epistemic merits of reflection. Denying that such a re-evaluation is called for would have been unnecessary, were the role of reflection in agency taken to be independent of the epistemic merits of the former.

<sup>6</sup> It is not completely clear on Sosa’s account whether the “understanding” in question needs to be a result of conscious reflection, or whether a certain counterfactual sensitivity to contrary evidence is sufficient. Some passages seem to suggest the latter (e.g., Sosa 1991, p. 240 and 2007, p. 111), others the former (e.g., 2007, p. 117). See Kornblith (forthcoming) for a discussion.



However, if Sosa is *not* right in assuming that we in general are better off for reflecting as opposed to not reflecting, then it would seem that, although epistemic agents may still express their agency by reflecting, we lack good reason to endow reflection with a particularly central role, at the expense of other cognitive tools or strategies with perhaps more impressive epistemic credentials.

How do we determine the epistemic merits of reflection? According to Sosa, cognitive faculties are *field-specific*.<sup>7</sup> In other words, what is relevant for the epistemic evaluation of a faculty is not simply reliability, but reliability with respect to a certain set of propositions. And how are such fields individuated?

Just how fields are defined is determined by the lay of interesting, illuminating generalizations about human cognition, which psychology and cognitive science are supposed in time to uncover. [...] Human intellectual virtues are abilities to attain cognitive accomplishments in “natural” fields, which would stand out by their place in useful, illuminating generalizations about human cognition (Sosa 1991, p. 236).

In other words, if we want to know how to individuate faculties for the purpose of evaluating the reliability of those faculties, we need to look to cognitive science. So are we, as a matter of empirical fact, in general better off from an epistemic point of view for reflecting?

Relevant experimental evidence suggests that the answer is “no.” For one thing, it is well known that we suffer from a variety of cognitive biases when reasoning under uncertainty.<sup>8</sup> This is not a problem for reflection *per se*, since many of the biases in question pertain to non-reflective cognition. But keep in mind that the idea is supposed to

---

<sup>7</sup> See Sosa (1991, p. 236).

<sup>8</sup> See Gilovich *et al.* (2002).

be that, by reflecting on our first-order beliefs, formed non-reflectively, we will be made better off from an epistemic point of view, due to the corrective influence of reflection. Such an optimistic take on the power of reflection simply does not hold up in light of the evidence, one important reason being that we systematically underestimate the extent to which we (as opposed to everyone else) suffer from cognitive bias.<sup>9</sup> More specifically, when reflecting on our grounds for belief, we have a strong tendency to conclude that our beliefs are well supported, even when they are not.<sup>10</sup> For this reason, reflecting on our grounds for belief does not invariably or even in general make us epistemically better off than not reflecting.

In fact, we are sometimes *worse* off for reflecting, as opposed to not reflecting. The last fifty years of predictive modeling suggest that we may often increase our reliability significantly by rejecting labor intensive strategies that intuitively strike us as epistemically responsible—such as collecting large quantities of data, weighing all available evidence, and defecting from general rules in so far as the situation is perceived to be relevantly unique—and instead opting for often extremely simple predictive models.<sup>11</sup> In fact, the greatest challenge for those developing such models is no longer to identify reliable models, but to convince people that they will be better off for using them. In other words, it is not lack of reliability but defection from policies urging that we use the models in question that is the main challenge. Moreover, as we will discuss at length in chapter 5, one of the primary activities driving up the defection rates is subjects

---

<sup>9</sup> See Pronin *et al.* (2002).

<sup>10</sup> This finding should be considered against the background of a large body of experimental research suggesting that we have a general tendency to overestimate our abilities (cognitive or otherwise). See Taylor (1989) for an overview of relevant research.

<sup>11</sup> See Dawes *et al.* (2002).

reflecting on their reliability and coming to believe—falsely and as a result of overconfidence—that they will be better off by not relying on the models in question.<sup>12</sup> In other words, reflecting on our grounds for belief sometimes makes us epistemically worse off than we would have been, had we opted for strategies that do not involve reflection.

What is the implication for how we ought to think about the relation between reflection and epistemic agency? It is *not* that reflection is a useless intellectual tool, or that reflecting always leaves us worse off than not reflecting. Nor is it that decisions about when or on what to reflect may not constitute expressions of epistemic agency. Reflection has just turned out to have a mixed epistemic track-record—like most of our intellectual tools—and there is, as such, no reason to consider it to be more closely related to epistemic agency than is any other goal-directed attempt to form true belief. In short, from the point of view of agency, reflection is not special.

This gets back to a question raised above: Is reflecting the *only* action relevant to epistemic agency? It will be suggested below that the answer is “no,” and if we want to understand what agents can actually do to increase their chances for forming true belief, and, thereby, identify the ways in which questions about agency may be relevant to questions about amelioration, we need to broaden our understanding of the domain of agency, rather than focus exclusively on reflection.

---

<sup>12</sup> See Sieck and Arkes (2005).

### 1.5. Is Epistemic Agency a Myth?

In a recent investigation into some of the same philosophers discussed above, Hilary Kornblith (forthcoming) expresses some skepticism about whether there really is a legitimate notion of epistemic agency to be found. Concluding his discussion of Sosa's take on the connection between reflection and agency, Kornblith writes that "the appeal to epistemic agency seems to be nothing more than a bit of mythology" (p. 118). About Moran's account of rational authority, Kornblith suggests that the considerations introduced by Moran "do not provide us with convincing reasons to believe that there really is such a thing as epistemic agency" (p. 126). As made clear above, there are certainly things to be said about the particular ways in which Sosa and Moran are characterizing epistemic agency, at least in so far as it pertains to epistemic amelioration. Still, Kornblith's skeptical conclusion regarding the very phenomenon of epistemic agency seems premature.

The reason why this is so has been hinted at several times above, but let us revisit the relevant issues in the context of Kornblith's complaints. One of his complaints is that moving from the unreflective to the reflective level of belief-formation does not introduce any epistemic agency; in both cases, what beliefs I form is not up to me. Kornblith makes his point in relation to an imagined scenario in which he is serving on a jury. While serving on a jury, you may be doing several things, such as focusing your attention on various pieces of evidence, as well as posing questions about their probity. Focusing your attention thus, however, is no different as far as agency is concerned from you turning your head or opening your eyes when forming various perceptual beliefs. Kornblith writes:

Whether I turn my head is determined by my choice, but once my head is turned in a certain direction, with my eyes open, and the lighting just so, my perceptual mechanisms will simply operate in me in ways which have nothing at all to do with the fact that I am an agent. The fact that I focus my attention, and question the relevance and probity of the evidence, thus show no more agency when I reflect than goes on in unreflective cases. Indeed, these activities not only show no more epistemic agency than goes on in unreflective cases in human beings; they show no more epistemic agency than goes on in lower animals when they form perceptual beliefs. But this is just to say that these features of reflectively formed belief do not exhibit epistemic agency at all (Kornblith forthcoming, p. 117).

The presupposition here seems to be that there is no *epistemic* agency if there is no *doxastic* agency, i.e., agency with respect to what beliefs to form. Only under this presupposition does it follow from both the reflective and the unreflective case involving an *element* of non-agency, in that *what* beliefs we form in response to the relevant stimuli is not within our voluntary control, that neither case involves *any* agency whatsoever. However, there is a perfectly good sense in which both the reflective and unreflective scenario may involve a kind of epistemic agency. The best argument to this effect is that only by acknowledging that this is so can we distinguish clearly between the kind of ameliorative advice that is bound to be unhelpful, and the kind of advice that may be effectively prescribed.

Consider the reflective case first. As we have seen, someone reflecting on her beliefs does not, thereby, gain any control over what beliefs she form as a result of reflecting. Consequently, any attempt at improving her reasoning by telling her what beliefs to form or not to form on reflection are bound to be completely unhelpful. That said, there are many other kinds of advice that she may be given, e.g., with respect to what kind of evidence-gathering practices she should involve herself in over others,

under what conditions she is better off subjecting her beliefs to reflective scrutiny as opposed to opting for strategies that involve no such scrutiny, and so on. Providing such advice is worthwhile exactly because what she does on these scores is within her control. Moreover, in so far as she chooses one option over the other in a manner that can be properly characterized as directed at the goal of forming true belief, she is expressing her epistemic agency. And in so far as she makes choices—perhaps, by heeding to sound ameliorative advice on the issue—that will actually make her better off with respect to the formation of true belief, she is expressing her epistemic agency wisely.

Next, consider the unreflective case. In particular, consider a medical student in the process of learning to read an x-ray plate. While no one has direct influence over *what* perceptual beliefs they form as a result of observation, the medical student may, clearly, benefit from being told exactly *how* and *where* to look at the plate. Advising her thus is worthwhile exactly because what she chooses to do here is within her control. Moreover, in so far as she looks in the right way and on the right locations, and, eventually, arrives at an accurate diagnosis, she can be said to not only be expressing her epistemic agency but to be expressing it wisely. And that is so irrespective of whether her attention at any point in the process of following her teacher's instructions moves from the x-ray plate to a second-order reflection on her first-order perceptual beliefs. As previous sections have made clear, what is crucial as far as the expression of epistemic agency is concerned is *not* that the agent in question reflects on her beliefs, but that she is doing things (reflectively or not) in a manner that can be characterized as directed at the goal of forming true belief.

Kornblith may, of course, grant this point, while maintaining that *this* kind of epistemic agency is not the kind of agency that the philosophers traditionally writing on the topic are interested in. On this point, consider what Kornblith has to say about the passage from Moran quoted above (on p. 26), where he elaborates on the relation between expressing rational agency, on the one hand, and what it means to (as he says) hand over one's belief to the authority of reason. According to Kornblith, there is a tension between these two claims—a tension that was called attention to above in relation to the relevant quote. According to Kornblith, the tension is such that we have reason to give up the former claim (about agency), while accepting a qualified version of the latter claim (about the authority of reason). However, as called attention to above, the tension only arises if we take the relevant authority to require an ability to opt out of the demands of rationality and, in effect, exert control (as opposed to mere authority) over what beliefs we form as a result of deliberating. As far as Moran's account of authority is concerned, no such thing is required. The only thing that is required is that we leave it up to our reflective faculties to lead us toward the proper conclusion, under the assumption that they are sensitive to the demands that rationality places on us.

At this point, Kornblith may suggest that Moran is no longer talking about agency in any robust sense of that term, and here me and Kornblith may be in agreement—at least if the relevantly robust sense of agency is one that can figure constructively in ameliorative contexts, in ways that I argued above that Moran's account of authority cannot. But as we have seen, this does not suggest that there is no such thing as epistemic agency; at most, it goes to show that there are theories about agency that fail to account

for one important role that a concept of epistemic agency should play, i.e., exactly the role that the present study attempts to delineate.

The same goes for Kornblith's final complaint about epistemic agency, i.e., that the idea of epistemic agency is wedded to an overly optimistic view about the epistemic merits of reflection. Here, we are in agreement about what the relevant empirical research shows. But as argued above, showing that reflection cannot carry the epistemic weight that some philosophers (Sosa, supposedly, being one of them) has wanted to put on its shoulders, neither shows that the decisions that go into when and on what to reflect may not constitute expressions of epistemic agency, nor suggests that there is no such thing as agency. If we acknowledge the weak notion of epistemic agency relied upon here, not even evidence to the effect that reflecting does not generally improve our reliability suggests that we may not express our agency by reflecting. At most, it suggests that we are in need of amelioration with respect to an epistemic activity that we perhaps did not think would require improvement, and that it will serve us well as epistemic agents to consider alternative ways to express our agency, in the event that the mechanisms responsible for the epistemic shortcomings of reflection prove recalcitrant enough.

## 1.6. Conclusion

Where does what we have seen above leave epistemic agency? Exactly where we left it. Agency still pertains to actions performed in an effort to attain our epistemic goals and, in particular, the goal of believing truly. It just turned out that (a) epistemic agency cannot pertain to decisions about what to believe, since making such decisions is impossible; (b)



if there is a sense in which what beliefs we end up with as a result of reflection is up to us, this does not serve to delimit a notion of agency that is relevant to amelioration; and (c) reflection, due to its largely mixed epistemic track-record, should neither be considered central to agency, nor, consequently, taken to be essential to the endeavors associated with ameliorative epistemology.

Keeping these points in mind, we see that there is a sense in which the question of what actions constitute epistemic agency is almost trivial. After all, what is it to engage in actions directed at the goal of forming true belief? It is to engage in *inquiry*. In other words, the actions by way of which we express our epistemic agency are those we perform when conducting inquiry. When conducting inquiry, we sometimes reflect on our first-order beliefs. But we do so much more than that—and rightfully so, given what we have found about the ways in which reflecting sometimes leads us astray due to a combination of bias and overconfidence. We also gather information, choose among different methods of investigation, and so on. And in so doing, we are expressing our epistemic agency.

At the same time, the very triviality of this answer to the question of what actions are relevant to epistemic agency suggests that the interesting questions regarding such agency lie elsewhere. One question pertains to an assumption that was made at the beginning of this chapter and relied on throughout, namely that believing truly is our only epistemic goal and, consequently, also unique in being of non-instrumental epistemic value. This assumption will be elaborated on as well as defended in the next two chapters. Another question regarding epistemic agency is an ameliorative one, namely: How exactly are we to bring the ways in which we actually express our epistemic agency

in line with the ways in which we should express our epistemic agency, given the goal of believing truly? This question will be postponed until chapter 5.

## CHAPTER 2

### I CAN'T MAKE YOU LOVE TRUTH

#### (AND YOU CAN'T MAKE ME LOVE EVIDENCE)

##### 2.1. Goals and Values

Why take true belief to be epistemically valuable irrespective of whether believing truly is conducive to something else of epistemic value? In short, why take true belief to be of *non-instrumental* epistemic value? One straightforward answer invokes the idea that true belief is a goal of inquiry or cognition.<sup>1</sup> More specifically, the relevant line of reasoning can be reconstructed as follows:

- (1) True belief is a goal of inquiry.
- (2) If something is a goal of inquiry, it is of non-instrumental epistemic value.
- (3) Hence, true belief is of non-instrumental epistemic value.

Before turning to (1), consider the rationale for (2). It seems reasonable to believe that goals of inquiry determine what states, processes, practices, and so on, are epistemically valuable. Not in the sense of what is *actually* valued—people fail to value what is valuable for all kinds of psychological reasons of no direct relevance to axiology—but rather in the sense of what is worthy of pursuit, given participation in practices that can

---

<sup>1</sup> See, e.g., Lynch (2009*a*), Alston (2005, p. 30), Haack (1993, p. 199), and BonJour (1985, pp. 83-84) for four representative statements, and David (2001) for further references.

be plausibly characterized as aiming toward the goals in question. More specifically, if true belief is a goal of inquiry, then true belief is of non-instrumental epistemic value and everything that is conducive to true belief is (at least) of *instrumental* epistemic value, in virtue of being conducive to the attainment of a goal of inquiry.

Someone might deny (2) by maintaining that goals of inquiry are one thing and epistemic value another, perhaps on account of the latter being a function of the fulfillment of epistemic obligations rather than of the attainment of goals of inquiry.<sup>2</sup> In other words, someone may argue that epistemic value is to be measured by something more akin to *effort* than to actual *achievement*, perhaps in a manner mirroring the distinction in ethics between non-consequentialist and consequentialist accounts of moral rightness.

There might be theoretical room in epistemology for effort-based conceptions of epistemic value. For present purposes, however, we may disregard such conceptions, the reason being that we are concerned with amelioration. Since the purpose of amelioration is to say something constructive about how to improve our epistemic ways in the contexts where we are failing to attain our epistemic goals, the kind of success we are concerned with in amelioration is achievement- rather than effort-based success. Effort, of any kind, is only of interest to the extent that investing great effort in achieving one's epistemic goals is correlated with actual achievement.<sup>3</sup>

---

<sup>2</sup> Feldman (1988) comes closest to defending such a position. As we shall see below (§2.2 and n. 14), however, he no longer challenges (3) by denying (2), but by denying (1).

<sup>3</sup> Interestingly enough, empirical evidence suggests that there is no strong correlation between a great effort put into living up to the traditional canons of epistemic obligation on the one hand, and the formation of true belief on the other. In fact, in some contexts, the correlation is even negative. See Bishop (2000).

Returning to (1), and setting aside the questions of whether (a) true belief is the *only* goal of inquiry, and (b) the goal is to be restricted to truths that are in some relevant sense *significant*—questions that will be addressed in chapters 3 and 4, respectively—why accept the claim that true belief is a goal of inquiry? William Alston finds himself at a loss for words:

I don't know how to prove that *the acquisition, retention, and use of true beliefs about matters that are of interest and/or importance* is the most basic and central goal of cognition. I don't know anything that is more obvious from which it could be derived (Alston 2005, p. 30; emphasis in original).

Many epistemologists side with Alston here.<sup>4</sup> For ease of reference, let us refer to any philosopher who takes true belief to be non-instrumentally valuable as a *veritist*.<sup>5</sup> As we shall see in §2.2, however, there are epistemologists who do not consider obvious what Alston does. Richard Feldman, for one, blocks the inference to (3) by denying (1). The goal of inquiry, he maintains, is to fulfill one's epistemic obligations by believing in accordance with one's evidence, where true or reliably formed belief is neither necessary nor sufficient for believing thus. What is particularly interesting about his position is that Feldman, thereby, rejects (3), while steering clear of two problematic claims, successfully harpooned in the literature.

The first claim is that true belief is of no value *whatsoever*.<sup>6</sup> Feldman is committed to no such claim. For one thing, he may reject (3) while consistently maintaining that believing truly may be of *instrumental* epistemic value, to the extent

---

<sup>4</sup> See references in n. 1.

<sup>5</sup> The term is borrowed from Goldman (1999), who introduces the notion of veritism to capture the particular brand of epistemology that is concerned with the varieties of processes and practices involved in the production of true belief.

<sup>6</sup> Stich (1990) comes closest to defending such a claim.

that believing truly is conducive to believing on the basis of sufficient evidence. For another, he may also maintain that true belief is of instrumental *pragmatic* value, in virtue of how accurate plans for action facilitate desire-satisfaction,<sup>7</sup> or of instrumental *moral* value, given how having an informed citizenry is conducive to the smooth working of democratic governance,<sup>8</sup> to name but two possibilities.

The second claim is that there are no constraints on what one's epistemic goals could be, nor, consequently, anything that stops one from taking virtually any good to be of non-instrumental epistemic value.<sup>9</sup> For reasons that will be discussed in §2.3, Feldman neither makes nor needs to make any such claim, and the arguments that can be leveraged against the claim in question, while establishing that there are *some* substantive constraints on what one's epistemic goals could be, neither single out true belief as the only candidate, nor rule out believing upon the basis of sufficient evidence as a goal of inquiry. Or so it will be argued.

Does it follow that we should give up on the idea that true belief, as opposed to, believing upon sufficient evidence, is non-instrumentally valuable? No. As we shall see below, the dialectical situation between the evidentialist and the veritist is symmetrical, in that neither can impose their respective notion of value on the other—a conclusion that will be drawn from the failure of several attempts on the part of both the evidentialist and the veritist to break this symmetry, discussed in §§2.3-5 below. It is concluded in §2.6 that, while there is no need to reject the evidentialist's take on value in order to protect the idea that truth is of non-instrumental value (although not necessarily for

---

<sup>7</sup> See Kornblith (2002).

<sup>8</sup> See Lynch (2004, pp. 174-176).

<sup>9</sup> See Field (2001) for an endorsement of this claim, and Lynch (2009b) for a critique.

*everyone*) from the evidentialist's contention that it is not, the prospects for convincing someone who does not share one's love for the relevant non-instrumental epistemic goods to start doing so seem dim indeed.

## 2.2. Feldman on Epistemic Value

According to Feldman (2002), "epistemological success amounts to having justified cognitive attitudes," which, in turn, "amounts to following one's evidence" (p. 382).

Does this claim conflict with the idea that true belief is of non-instrumental epistemic value? That depends on how we spell out what it is to *follow* one's evidence, which in turn needs to be analyzed in terms of *having* evidence. On one understanding of what it is to have evidence, something is evidence for something else if and only if the former is as reliable indicator of the latter.<sup>10</sup> That, however, is not how Feldman understands what it is for something to be evidence, the primary reason being that he wants it to be possible for someone in an evil-demon scenario to be justified and, consequently, also to have evidence, despite being subject to massively misleading experiences.<sup>11</sup> In an evil-demon scenario, whatever evidence a subject has in virtue of undergoing certain experiences, these experiences do *not* reliably indicate anything about the subject's surrounding.

What is it, then, about evidence that justifies belief, if not a relation of truth-indication? According to Feldman—here, writing with his long-time partner in evidentialism, Earl Conee—evidence justifies in virtue of a certain *coherence* relation:

---

<sup>10</sup> See, e.g., Goldman (forthcoming).

<sup>11</sup> See Feldman (1985).

[...] the general idea is that a person has a set of experiences, including perceptual experiences, memorial experiences, and so on. What is justified for the person includes propositions that are part of the best explanation of those experiences available to the person. [...] The best available explanation of one's evidence is a body of propositions about the world and one's place in it that makes best sense of the existence of one's evidence. This notion of making sense of one's evidence can be equally well described as fitting the presence of the evidence into a coherent view of one's situation. So it may be helpful to think of our view as a non-traditional version of coherentism. The coherence that justifies holds among propositions that assert the existence of the non-doxastic states that constitute one's ultimate evidence and propositions that offer an optimal available explanation of the existence of that evidence (Conee and Feldman, 2008, p. 98).

In other words, a person's (ultimate) evidence consists in a set of experiences, and sets of propositions are justified in so far as they provide explanations of those experiences that cohere with propositions asserting the presence of the experiences in question.

Coherence is a complex notion with a long history, and Feldman and Conee, unfortunately, do not spell out the details of the particular notion that they have in mind. However, there is an emerging consensus in the literature that coherence is not truth-conducive, meaning that more coherence does not imply a higher likelihood of truth, even *ceteris paribus*.<sup>12</sup> More than that, such a disconnect between coherence and truth would seem to be exactly what Feldman needs in order to accommodate the idea that someone in an evil-demon scenario could be fully justified despite undergoing massively misleading experiences, as long as the propositions believed by the subject in question provide explanations of her experiences that cohere with the presence of those experiences.

This is not to subscribe Feldman to a view on which there is *no* connection between following one's evidence and believing truly. For one thing, Feldman takes it

---

<sup>12</sup> See Olsson (2005).



that we can make a good case for most justified beliefs being true in the *actual* world.<sup>13</sup> In other words, Feldman's position is compatible with it being the case that, in a world like ours, where experiences as a matter of contingent fact tend to provide pretty good information about the external world, basing one's beliefs upon one's evidence is a good means toward forming true belief. Crucially, however, Feldman denies that this is something that holds by necessity. Again, in an evil-demon scenario, no degree of diligence with respect to basing one's belief upon one's evidence will enable one to form true beliefs about the external world, which is why neither true belief nor reliability is necessary for justification nor, consequently, for epistemic success, according to Feldman, given his identification of the two.

Notice that a veritist may grant this. Nothing said so far implies that true belief is the *only* goal of inquiry. However, Feldman is not only skeptical about the necessity of truth or truth-conduciveness for epistemic success, but also about it being *sufficient*. Registering his skepticism about the idea that mere true belief is an epistemic goal, and attaining such belief, consequently, constitutes a form of epistemological success, Feldman poses the following hypothetical:

[...] suppose a person acquires strong evidence against a proposition he has long defended in public. Out of stubbornness, the person retains the old belief. And suppose that, contrary to the new evidence, the old belief is in fact true. If the goal is simply truth, he's achieved the goal and is, in this case at least, an epistemological success. But, as Locke said of a person who does not reason as he ought, "however he sometimes lights on truth, is in the right but by chance; and I know not whether the luckiness of the accident will excuse the irregularity of his proceeding." [...] Setting aside questions of excuses, the idea here seems to be that the person who reasons badly and stumbles onto a truth is not believing as he ought, is not achieving epistemological success (Feldman 2002, p. 378).

---

<sup>13</sup> See Feldman (2003, pp. 615-616).

Elaborating on a related scenario a couple of sentences further down, Feldman makes it clear that he sides with Locke on this point. To simply achieve true belief is *not* to achieve any goal of ours, nor to be any kind of epistemological success:

Imagine a person who makes an unreasonable and unreliable inference that happens to lead to a true belief on a particular occasion. It might be fortunate that he's got this true belief, but I see nothing epistemologically meritorious about it (Feldman 2002, p. 379).

Clearly, the idea that there is *nothing* epistemically meritorious about (mere) true belief—that is, that believing truly is neither *a* goal nor *one* kind of epistemic success—runs contrary to any account of epistemic value relying on argument (1)-(3) above.<sup>14</sup> Granted, few veritists would deny that there is *something* epistemically lacking about the person in the scenario Feldman imagines. For example, reliabilists about justification would deny that she is justified. However, what is at issue here is whether there is *anything* of epistemic value about her epistemic situation, and that is where the veritist parts company with Feldman. Is Feldman in any relevant sense, thereby, *mistaken* about epistemic value? The next section considers two arguments—both of which are, ultimately, rejected—to the effect that he is.

---

<sup>14</sup> Feldman (in correspondence) confirms that there has been a shift in his views on epistemic value, from his (1988), where he accepts that true belief is a goal of inquiry and simply denies that true belief has anything to do with value, to his (2002), where he (as we have seen) denies that true belief has anything to do with value by denying that it is a goal of inquiry.

### 2.3. Is Feldman Mistaken About Value?

Recently, Michael Lynch (2009*a* and *b*) has provided a series of arguments against what he calls *epistemic expressivism*, or the view that there are no constraints on what one's epistemic goals could be, nor, consequently, on what one may take to be of non-instrumental epistemic value—a position that he attributes to Hartry Field (2001).

Presently, we will not be concerned with epistemic expressivism. However, as we shall see, at least two of Lynch's arguments (if successful) apply equally well to someone who, like Feldman, simply denies that truth is of non-instrumental epistemic value.

First, consider what we may refer to as the *no inquiry argument*. The idea is that someone who is forming belief in a manner that involves no commitment to the idea that propositions are to be embraced if and only if they are true is not engaging in inquiry.<sup>15</sup>

Lynch (2009*a*) makes this point in relation to a hypothetical example including someone—let us call him Steve—who is committed to the goal of “accepting” all and only propositions that flatter his cultural origin. Is Steve, thereby, engaging in inquiry? Lynch suggests that the answer has to be “no,” since the practice Steve is involved in does not involve a commitment to the idea that propositions are to be embraced if and only if they are true. Whatever practice Steve is involved in, it is not inquiry.

When applied to Steve, this line of reasoning is highly convincing. But does the point in question generalize to *any* practice that does not involve a commitment to the idea that propositions are to be embraced if and only if they are true, including a practice

---

<sup>15</sup> The norm of embracing propositions if and only if they are true is, of course, not meant to be action-guiding. Instead, the norm is supposed to tell us something about conditions under which particular instances of inquiry have to be considered successful or unsuccessful, namely to the extent that they terminate in true belief or false belief, respectively.

that involves (nothing but) a commitment to the idea that propositions are to be embraced if and only if they are based upon sufficient evidence, as spelled out by Feldman? It is not so clear that it does. To see why, consider two ways in which the relevant point about what constitutes inquiry could be pushed. On the one hand, it could be maintained that the norm according to which we should believe something if and only if it is true attaches to the relevant notion of inquiry by *conceptual* necessity. In that case, the dispute seems to boil down to one about meanings, and the veritist would have to resort to either suggesting that Feldman is conceptually confused—a claim that is highly implausible in light of him being one of contemporary epistemology’s most prominent contributors—or granting that the evidentialist is using a different yet equally legitimate notion of inquiry.

On the other hand, Lynch makes it clear that this is not how he views matters:

[...] it isn’t at all obvious that the above arguments need to be understood as being about the meanings of “inquiry” or “belief” or “language”. As we have noted earlier, (TN) [i.e., the idea that it is *prima facie* correct to believe <p> if and only if <p> is true] and what follows from it needn’t be thought of as conceptual truths. One could take such points to be about inquiry and belief rather than “inquiry” and “belief”. Cooking aims to produce food; comedy aims to produce laughter. Likewise, one might think, inquiry aims to produce true beliefs. It is a fact about such practices that they cannot be separated from their aims; their aims are partly constitutive of them (Lynch 2009*b*, pp. 90-91).

In other words, rather than pertaining to conceptual necessities, the relevant truths about inquiry may, as Lynch suggests, be truths of *constitution*. However, at this point, the argument starts to look question begging, at least if leveraged against the evidentialist. More specifically, the argument against counting above evidentialist practice as a form of inquiry would, under this construal, require a premise to the effect that inquiry is the

pursuit of truth, which is exactly the claim that the evidentialist contests. Clearly, the evidentialist has no reason to grant the veritist that premise.<sup>16</sup>

Lynch seems to admit this point when noting that he “cannot give a non-circular justification of my belief that it is valuable to engage in inquiry [as construed by the veritist]; for in answering the question I am already committed to the value of the very practice in question” (2009a, p. 239). Moreover, he grants that, far-fetched examples like the one with Steve aside, it seems at least *prima facie* reasonable to say that someone may engage in inquiry in a context where the latter is spelled out in terms of the goal of having a coherent belief system, or having epistemically rational or justified beliefs. At this point, however, the second argument enters, an argument that we may refer to as the *presupposition argument*. Lynch writes:

No doubt, people can and do have such goals in pursuing inquiry. The question is whether [...] we can conceive of someone who has distinct epistemic goals that don't *include* the goal of having true belief. Of course, the typical way of understanding inquiry does require that it have only one goal: truth [...]. The thought here is that the value of achieving, e.g. a coherent belief system lies in the fact that doing so is instrumental to believing what is true and not believing the false. [...] If we didn't think this, it seems unlikely that we would care about whether our beliefs were coherent. But of course if this is how we see the matter, having the goal of coherence or justification presupposes having the goal of true belief (Lynch 2009b, p. 89).

The problem is, of course, that this is *not* how the evidentialist, at least of Feldman's stripe, sees the matter. As we saw above, he denies exactly that the value of justification,

---

<sup>16</sup> The same points made here about the futility of discrediting the evidentialist's account of value with reference to a constitutive or conceptual connection between truth and inquiry can be made, *mutatis mutandis*, against any analogous attempt in terms of there being a constitutive or conceptual connection between truth and *belief*, on account of the former being the norm of the latter, and inquiry being the practice of belief-formation. See, e.g., Lynch (2009a).

i.e., of basing your beliefs upon sufficient evidence, is parasitic upon the value of believing truly. Suggesting that the evidentialist's axiology presupposes such a relationship would be to misconstrue it, at best, and to beg the question, at worst. As such, we may conclude that neither of Lynch's arguments establishes that Feldman is mistaken about epistemic value.

It is crucial to note, however, that the dialectical situation is symmetrical, in that any evidentialist elaboration on the notion of inquiry is equally unlikely to sway the veritist, as above reflections on the veritistic notion are to sway the evidentialist—and rightly so. As such, nothing said so far goes to show that the veritist should stop valuing true belief non-instrumentally. The point is merely that neither the veritist nor the evidentialist can impose their respective notion of value on the other, as long as the relevant values are grounded in nothing but different conceptions of inquiry. The next section considers an argument on part of the veritist that, if successful, would break the symmetry, by grounding the value of true belief in the fact that certain true beliefs are necessary parts of a flourishing life. If successful, the argument would provide reasons for the evidentialist to start valuing (at least some) true beliefs non-instrumentally, irrespective of what happens to be her conception of inquiry.

#### 2.4. True Belief and Human Flourishing

According to Lynch (2004), we *care* about having true beliefs, where “caring for something entails that we treat it as [an] *end*” (p. 119). Differently put, we take true belief to be of non-instrumental epistemic value. But what *reasons* do we—as in: *all* of

us, irrespective of our particular notions of inquiry—have for valuing truth thus? Lynch provides two arguments meant to answer this question, both of which elaborate on the intimate relationship between believing truly and being happy, in the sense of leading a flourishing life.<sup>17</sup> He refers to the first argument as the *argument from self-knowledge*. Let us consider Lynch’s formulation of the argument, before turning to his definitions of the notions invoked (and I quote):

- (1) Self-respect and authenticity require a sense of self.
- (2) Because it is required for self-respect and authenticity, which are part of happiness, having a sense of self is an important part of happiness.
- (3) Having a sense of self means having true beliefs about what you care about.
- (4) Therefore, having some true beliefs—about what you care about—is also part of happiness, other things being equal (Lynch 2004, p. 127).

“If we grant,” Lynch continues, ”that whatever is part of happiness is worth caring about for its own sake, from these premises we can deduce that the truth about what you care about is itself worth caring about for its own sake” (p. 127). In other words, Lynch wants to infer the conclusion

- (5) the truth about what you care about is worth caring about for its own sake,

from (1)-(4), together with the following additional premise:

- (AP) If  $x$  is part of some  $y$  (e.g., happiness) that is worth caring about for its own sake, then  $x$  is worth caring about for its own sake.

---

<sup>17</sup> Actually, Lynch provides six arguments. However, three of these pertain to instrumental values (see n. 8), and the fourth fails for the same reason as the first and second (see n. 20), which leaves the two arguments to be discussed here.

Let us consider the premises in turn. To have *self-respect*, as Lynch understands it, is to have a sense of your own value (p. 124), while *authenticity* is to identify with the desires that effectively guide your actions (p. 125). To have a *sense of self* is to have true beliefs about what you care about, as per premise (3). Since it is, arguably, impossible to either have a sense of your own value, or to identify with the desires that effectively guide your action, without having true beliefs about what you care about, self-respect and authenticity require a sense of self. Hence, premise (1).

Premise (2) maintains that self-respect and authenticity are parts of happiness.

What does it mean to be a *part* of something here? Lynch explains:

*Being a part of something is crucially different from being a means to it. A means to an end is ipso facto different from the end itself. A necessary part of something, on the other hand, helps make the whole of which it is a part the thing that it is. Change or destroy the part and you change or destroy the whole (p. 128; emphasis in original).*

In other words, for some  $x$  to be a part of some  $y$  is for  $x$  to be a *necessary* part of  $y$ , such that the absence of  $x$  implies the absence of  $y$ . Interpreting premise (2) accordingly, and granting premise (3)—i.e., that having a sense of self means having true beliefs about what you care about—it follows that having true beliefs about what you care about is a necessary part of happiness. If that is how we are to understand what it is to be a part of something, (4) can be reformulated as follows:

(4\*) Having true beliefs about what you care about is a *necessary* part of happiness, other things being equal.



It might seem strange to talk about something being a necessary part of something, other things being equal. After all, to be a necessary part of something is to be a part of something by necessity, i.e., *irrespective* of whether other things are equal. However, what Lynch means to indicate by the qualification in terms of “other things being equal” is that the value in question is *pro tanto*, in that it may be overruled by other considerations and, hence, is not an all-things-considered good.<sup>18</sup> In other words, the conclusion that we are to draw from Lynch’s argument is the following:

(5\*) The truth about what you care about is worth caring about for its own sake, in the sense that having the true beliefs in question is *pro tanto* good in itself.

If Lynch’s argument succeeds in establishing (5\*), then we have found an argument to the effect that at least some true beliefs are (*pro tanto*) valuable in themselves, irrespective of what happens to be one’s particular notion of inquiry. The problem with the argument, however, is (AP). We may see more clearly what the problem is by reformulating (AP) in light of Lynch’s definition of what it is for something to be a part of something as follows:

(AP\*) If *x* is a *necessary* part of some *y* (e.g., happiness) that is worth caring about for its own sake—i.e., if *x* is such that its absence implies the

---

<sup>18</sup> See Lynch (2004, pp. 46-51). This interpretation is confirmed by Lynch (in correspondence).

absences of that of which it is a part—then  $x$  is worth caring about for its own sake.

Clearly, (AP\*) is not true. It does *not* follow from  $x$  being a necessary part of  $y$ , that if  $y$  is worth caring about for its own sake then the same goes for  $x$ . By way of illustration, assume that true belief is worth caring about for its own sake. If (AP\*) were true, it would follow that *mere* belief, being a necessary part of true belief, would be worth caring about for its own sake. This clearly cannot be right, even if we qualify the good involved as being a mere *pro tanto* good. Hence, (AP\*) cannot be true. And if (AP\*) is not true, then we do not get the conclusion that Lynch wants to draw from (1)-(4), namely (5\*).<sup>19</sup>

Lynch's second argument, or what he refers to as the *argument from integrity*, suffers from an analogous problem. This argument attempts to establish the more general conclusion that the very act of caring about truth as such is non-instrumentally valuable. The reason, Lynch argues, is that caring about truth as such is essential to *intellectual integrity*. To instantiate intellectual integrity

---

<sup>19</sup> Notice that we can run the same argument even if we interpret the relevant relation between true belief and happiness in terms of supervenience. After all, it does not follow from  $y$  being valuable in itself that, if  $x$  is part of the supervenience base of  $y$ , then  $x$  is also valuable in itself. By way of illustration, assume (if only for the sake of the argument) that being a person is valuable in itself. Assume, furthermore, that facts about person-hood supervene on neurological facts, and that I am the person I am in virtue of instantiating the particular neurological facts  $N_1$ - $N_{850}$ . If you change any of those facts, you change (if not destroy) the whole, i.e., me. (You might still have *a* person, of course, but arguably not a person that is *me*.) Then, take any neurological fact in this series, say,  $N_{46}$ . Is instantiating  $N_{46}$  valuable in itself? It is not—despite it being part of the supervenience base of something that (by hypothesis) is valuable in itself. In other words, whatever value attaches to the supervenient does not trickle down to the individual parts that make up the subvenient, in the manner that Lynch's argument requires.

[...] one must be open to having true belief *in general*, and pursuing truth *in general*, on those questions that come before you, whatever those may be. Caring about truth for its own sake, that is, being willing to both pursue it and be open to it, is therefore a necessary condition for intellectual integrity (p. 133).

According to Lynch, intellectual integrity is what he refers to as a *constitutive* good, or something worth caring about for its own sake. We will return to this claim momentarily, but let us first consider the argument that he sets out (and I quote):

- (1) If intellectual integrity is a constitutive good, then so is caring about truth as such.
- (2) Intellectual integrity is a constitutive good.
- (3) Therefore, caring about the truth as such is a constitutive good.
- (4) If caring about truth as such is a constitutive good, then truth as such is worth caring about for its own sake.
- (5) Therefore, truth as such is worth caring about for its own sake (Lynch 2004, p. 136).

This is a valid argument, but it is not sound. Consider the conditional in premise (1).

Why would it follow from intellectual integrity being a constitutive good, that caring about truth as such is also a constitutive good? Lynch suggests that “[c]aring about the truth as such is a constitutive good because it is essential to intellectual integrity, and therefore, to integrity proper” (p. 136). But, as already noted above in relation to (AP\*), it does not follow from  $x$  being a necessary condition for  $y$ , that, if  $y$  is worth caring about for its own sake, then the same goes for  $x$ . Consequently, we cannot infer (3) from (1) and (2), nor (5) from (3) and (4).<sup>20</sup>

Two things need to be noted about what should *not* be inferred from the failure of Lynch’s arguments. First, while the arguments in question failed to establish that some

---

<sup>20</sup> The same goes for Lynch’s (2004, p. 157) structurally identical argument about sincerity.

true beliefs are non-instrumentally valuable (*pro tanto* or otherwise) irrespective of your particular conception of inquiry, nothing said so far rules out the possibility that having certain true beliefs (e.g., about what you care about), or perhaps even caring about truth as such, is under many circumstances *conducive* to leading a flourishing life. However, that at most establishes that some instances of true belief are instrumentally valuable, not that they are non-instrumentally valuable. As such, it provides no reason for those who do not already value (any) true belief non-instrumentally to start doing so.

Second, in accordance with what was noticed above, the failure of Lynch's arguments does not go to show that those who already value true belief non-instrumentally should stop doing so. But in want of an account of the non-instrumental value of true belief that appeals to something beyond conceptions of inquiry that are not universally shared, the veritist must accept the fact that she can provide no reasons for those who do not already share the relevant notion of inquiry to start valuing true belief non-instrumentally. The next section considers the matter from the other side, by investigating the possibility that the evidentialist can provide the kind of argument that Lynch could not, i.e., an argument to the effect that we should value basing our beliefs upon sufficient evidence non-instrumentally, irrespective of what happens to be our particular conception of inquiry.

## 2.5. Epistemic Value and the Ethics of Belief

In his essay "The Ethics of Belief" (1866), W. K. Clifford famously argues that it is wrong to believe upon insufficient evidence. Clifford argues by way of examples in

which beliefs formed on the basis of insufficient evidence have dire consequences for others, but suggests that believing thus would have been wrong even if the consequences in question had not ensued. Moreover, Clifford takes it that the duty not to believe upon insufficient evidence extends beyond those engaged in decisions involving high stakes. Indeed, the duty is universal. As he famously sums up his position on the issue: “It is wrong, always, and everywhere, and for everyone, to believe anything upon insufficient evidence” (p. 346).

For Clifford, this synchronic duty not to believe upon insufficient evidence follows from a more fundamental, diachronic duty of inquiry, i.e., “the universal duty of questioning all that we believe” (p. 343). In questioning our beliefs, we attempt to find out whether or not they are based on sufficient evidence. If they are, we may continue holding them, knowing that they have been “fairly earned by investigation” (*ibid.*). If they are not, “we have to begin again at the beginning, and try to learn what the thing [under investigation] is and how it is to be dealt with—if indeed anything can be learnt about it” (*ibid.*). In other words, through persistent inquiry, we either uncover sufficient evidence one way or the other and adjust our belief accordingly, or find that there is no evidence either way and that suspension of belief, consequently, is the most appropriate attitude to take toward the relevant proposition.

It is as a fruit of such “scrupulous care and self-control” (p. 347) that we, on Clifford’s picture, should understand the value of believing upon sufficient evidence. To believe thus is to have satisfied the duty of inquiry by making sure that, for every proposition under consideration, that proposition is only believed if it is based on sufficient evidence. As suggested by the case of belief suspension in situations where it

turns out that nothing can be learnt about the object under investigation, believing upon sufficient evidence may not be the *only* (albeit, possibly, the only *happy*) way of satisfying the duty of inquiry.<sup>21</sup> Still, if Clifford is right, even those who do not happen to share Feldman's conception of inquiry ought to value having their beliefs be based upon sufficient evidence non-instrumentally. To see why, consider the fact that Clifford, unlike the great majority of contemporary epistemologists, seems to take our epistemic duties to be a mere subset of our ethical duties.<sup>22</sup> With this in mind, consider the following argument:

- (1) We have a duty to only believe upon sufficient evidence, and the duty is an ethical one.
- (2) To believe propositions upon sufficient evidence is to satisfy the duty to only believe upon sufficient evidence with respect to the relevant propositions.
- (3) Satisfying an ethical duty is non-instrumentally valuable.
- (4) Hence, to believe upon sufficient evidence is non-instrumentally valuable.
- (5) The relevant ethical duty applies universally.
- (6) To believe upon sufficient evidence is non-instrumentally valuable for everyone.

---

<sup>21</sup> However, given Clifford's emphasis on our diachronic duty to question all that we believe and, in effect, inquire as far as possible, it is unlikely that he would consider complete suspension of belief as an acceptable way to satisfy the duty of inquiry, unless such suspension constitutes the inescapable end point of pain-staking investigation.

<sup>22</sup> See Haack (2001) for a convincing set of arguments to this effect.

Premise (1) captures the duty Clifford infers from the duty of inquiry, and identifies it as an ethical one. Premise (2) corresponds to aforementioned observation about how we may satisfy the duty in question by believing upon sufficient evidence. Premise (3) amounts to an assumption that will be granted for the sake of the argument, i.e., that satisfying our ethical duties—whatever they are—is non-instrumentally valuable. Moreover, if all epistemic duties are ethical duties, the same applies to the former, as per premise (4). Assuming that the relevant ethical duties apply universally, as per premise (5), believing upon sufficient evidence is non-instrumentally valuable not only for those who, like Feldman, share the evidentialist conception of epistemic success, but to everyone, as per (6).

Albeit a valid argument, premise (1) is highly implausible. To see why, consider that there, clearly, are cases in which one is *not* morally reproachable for believing upon insufficient evidence, as famously pointed out by William James (1907) in his critique of Clifford. Clifford denied this possibility, by suggesting that no belief is (morally) inconsequential.<sup>23</sup> In so doing, Clifford assumed that mere potential harm, however remote, is sufficient for moral culpability—a position that, as Susan Haack (2001) points out, would imply that “not only drunk driving, but owning a car would be morally culpable” (p. 27). Moreover, Clifford maintains that believing upon insufficient evidence invariably strengthens the habit of credulity, in turn amounting to a risk for harm.<sup>24</sup> Granted, belief upon insufficient evidence is sometimes the result of cognitive tendencies that may lead to harm. But in maintaining that believing upon insufficient

---

<sup>23</sup> See Clifford (1866, p. 342).

<sup>24</sup> See Clifford (1866, pp. 344-346).

evidence *invariably* leads to credulity, Clifford seems to assume that such belief *always* is the result of potentially harmful cognitive traits, and that is highly questionable.

Consequently, it would seem that Clifford has to grant James' point that there are cases in which believing upon insufficient evidence is morally permissible. In fact, contrary to Clifford's exaggerated sense of potential harm, a great many beliefs, including some held and debated within science and philosophy, are of no great moral consequence. If Clifford were to acknowledge this, and at the same time hold on to (1), he would have to maintain that the cases in question fall outside the scope not only of moral evaluation but (consequently) also of epistemic evaluation, the latter being a special case of the former. From this, it would follow that none of the beliefs in question are held in ways that are epistemically impermissible. Such epistemic *laissez-faire* seems undesirable, to say the least, and the way to block the conclusion is, of course, to deny (1), and maintain that, in so far as we have a duty to only believe upon sufficient evidence, that duty is not an ethical one.

In light of this, it should come as no surprise that contemporary epistemologists sympathetic to the idea of an ethic of belief, nevertheless, tend to make a distinction between ethical and epistemic duties. For example, in spelling out his evidentialist take on Clifford's position, Feldman (1988, pp. 236-237) explicitly rejects the identification of the domains of epistemic and moral obligations, and maintains that beliefs with morally bad consequences may still be epistemically obligatory. Similarly, when claiming that every person is required to try her best to bring it about that, for every proposition considered, she accepts the proposition only if it is true, Roderick Chisholm



(1977, p. 14) specifies the requirement as a *purely intellectual* requirement.<sup>25</sup> In other words, the main modern defenders of Clifford's project construes it not so much an *ethics* of belief as simply an *epistemology* of belief formation.<sup>26</sup>

In so doing, however, the neo-Cliffordians also surrender the one aspect of Clifford's theory that would provide them with a way to convince the veritist of the non-instrumental value of basing one's belief upon sufficient evidence. As argument (1)-(6) above makes clear, the reason Clifford's account was promising on this score was exactly because it took our duty to base our beliefs thus to be an ethical duty—a fact that, given the wide scope and non-instrumental nature of our ethical duties, would provide a lever for extending the normative force of the evidentialists' emphasis on the value of believing upon sufficient evidence beyond those who happen to share their particular conception of inquiry. In so far as this ambition is not present in modern defenders of an ethics of belief, they may thereby be constructing far more plausible accounts of the relevant evidentialist norms than Clifford did, albeit at the expense of providing little by way of arguments for convincing the veritist of the non-instrumental value of believing upon the basis of sufficient evidence.

---

<sup>25</sup> Chisholm does at times write as if epistemic duties are a mere subspecies of ethical duties, e.g., in his (1991, p. 119). However, when spelled out, the claim turns out to be that (a) ethical duties are duties that are not overridden by any other requirement, and (b) non-overridden epistemic duties (like any non-overridden duties), hence, are ethical duties (see pp. 127-128). Granting this somewhat questionable definition of an ethical duty, it still only goes to show that *some* epistemic duties are ethical duties, not that all are, nor consequently that the former are a mere subspecies of the latter.

<sup>26</sup> Wolterstorff (2005) is a possible exception.

## 2.6. Conclusion

The failure of surveyed accounts to break the dialectical symmetry between the veritist and the evidentialist does not prove that the symmetry cannot be broken. Still, the absence of further, more successful accounts leaves us with no reason to think that the requisite arguments are forthcoming. This is not to deny that we all have ample *instrumental* reason to value true belief, given the ways in which believing truly may further our pragmatic, moral or prudential endeavors. Furthermore, and as has been stressed several times already, what has been argued here is neither to deny that many of us *do* value at least some true beliefs irrespective of whether doing so is conducive to something else of epistemic value, nor to maintain that those of us who do should stop doing so. As we have seen, there is a perfectly good answer to the question of why true belief is of non-instrumental epistemic valuable, namely the one provided in terms of true belief being a goal of inquiry. It just happens to be an answer that cannot be invoked in an argument to the effect that those who do not share the relevant notion of inquiry should start valuing true belief non-instrumentally.

This fact, taken together with the failure of surveyed accounts to ground the relevant claims about value in something other than our notions of inquiry, suggests that the prospects for convincing someone who does not share one's love for the relevant non-instrumental, epistemic goods to start doing so seem dim indeed. However, given that failure to convince someone to value something by no means is a sufficient reason for ceasing to value that thing, there is no need to reject the evidentialist's take on value in order to protect the idea that truth is of non-instrumental value from the evidentialist's contention that it is not. Granted, I can't make you love truth, if you don't already do.

But by aforementioned symmetry you also can't make me love evidence. For this reason, subsequent chapters will assume the position taken by most epistemologists, to the effect that true belief is, indeed, a non-instrumental, epistemic good—albeit, as we have seen, not necessarily for everyone.

## CHAPTER 3

### IN DEFENSE OF EPISTEMIC VALUE MONISM

#### 3.1. Varieties of Value

As noted in the previous chapter, most epistemologists take believing truly to be of non-instrumental epistemic value in virtue of true belief being a goal of inquiry. It is far more controversial, however, whether the following thesis holds—a thesis that we may refer to as *epistemic value monism*, or EVM for short:

(EVM) Believing truly is unique in possessing fundamental, non-instrumental epistemic value.<sup>1</sup>

The present chapter defends EVM against a series of objections coming out of recent discussions about the value of knowledge. There is no doubt that these discussions have been extremely rewarding in bringing about something of an axiological renaissance in epistemology. Still, it is the thesis of the present chapter that recent pluralistic critiques of EVM fail to provide any convincing counterexamples, and that epistemic value monism remains a perfectly defensible thesis.

Before proceeding to EVM's defense, let us clarify its component notions. First, what is *epistemic value*? As noted in the previous chapter, it seems reasonable to believe that (a) goals of inquiry determine what activities, states, processes, practices, and so on,

---

<sup>1</sup> Recent proponents of EVM include Goldman (1999) and David (2005).

are epistemically valuable, (b) true belief is one such goal, (c) true belief, hence, is of non-instrumental epistemic value, and (d) everything that is an effective means to believing truly is of *instrumental* epistemic value. The distinction between instrumental and non-instrumental value pertains to how value may be inherited across causal relations. Another way in which value may be inherited is from parts to wholes. As we shall see in §3.2, that some states inherit value from one of their components is important to keep in mind when thinking about the comparative value of complex states. With this in mind, let us say that something is of *fundamental* value if and only if its value does *not* derive from the value of any of its components.

By way of illustration, EVM is compatible with knowing being of non-instrumental epistemic value in virtue of believing truly being a component of knowing.<sup>2</sup> What is denied is simply that such a state—or, more generally, any state that involves true belief as a mere component—is of fundamental, as opposed to derived, non-instrumental value. The notions of fundamental versus derived value may be applied also to instrumental values. However, it will suffice to make the distinction in relation to non-instrumental value for the purposes of the present investigation. Consequently, any unqualified use of the terms ‘fundamental’ or ‘derived’ shall, henceforth, be understood in terms of fundamental or derived *non-instrumental* value.<sup>3</sup>

---

<sup>2</sup> The relevant states do not literally have components. The components in question correspond to the conditions included in a correct analysis of the state(s) in question. If a state has no analysis—as Williamson (2000) has argued is the case for knowledge—it has no components, but may still be of fundamental value.

<sup>3</sup> The taxonomy developed here is not meant to cover all possible mechanisms for value inheritance. For example, Goldman (Goldman and Olsson, 2009) argues that there are instances in which the non-instrumental value of true belief may be inherited by token processes in virtue of being instances of reliable types—a pattern of inheritance that can be categorized neither as an instance of instrumental value nor as one of derived value.

It is controversial whether all bearers of non-instrumental value are *intrinsically* valuable, or valuable in themselves. For present purposes, two issues need to be addressed. The first issue is whether to understand intrinsic value as applying to objects or to states of affairs. If the former, it is easy to come up with examples involving objects identical in intrinsic properties, yet (intuitively) differing in intrinsic value.<sup>4</sup> However, if we take intrinsic value to apply to states of affairs, as is common in the Moorean tradition, such examples can be accounted for, and the claim that non-instrumental and intrinsic values coincide be maintained.<sup>5</sup> It is easy to see why one might want to opt for an object-construal of intrinsic value when it comes to a bearer of value that has figured prominently in the Kantian tradition, *viz.* the person. However, as far as epistemic value is concerned, it is reasonable to think that the bearers of value are states of affairs—be it states of affairs involving doxastic states, processes, practices, or what have you—rather than objects. One strong indication of this is that it seems odd to say that *a true belief* (an object) is valuable, unless this is just a different way of saying that *someone believing truly* (a state of affairs) is valuable, and similarly for someone knowing, understanding, being justified, and so on.<sup>6</sup> Construed thus, the notion of non-instrumental value, as understood here, may be taken to coincide with intrinsic value.

---

That said, a lack of taxonomical exhaustiveness would only be a problem if the taxonomy used somehow stacked the dialectical deck against the pluralist, which it does not. If anything, by not incorporating the kind of value that Goldman introduces and argues can be invoked in a solution to problems typically leveraged against monists, it might even be that the present taxonomy stacks the deck against the monist—which, of course, makes any defense of monism in terms of this taxonomy all the more significant.

<sup>4</sup> See, e.g., Rabinowicz and Rønnow-Rasmussen (2000).

<sup>5</sup> See Bradley (2006).

<sup>6</sup> Similarly, in so far as references will be made below to “true belief” or “knowledge” being epistemically valuable, this is not to suggest that *abstract* objects are proper bearers

Still, the discussions below will be framed in terms of non-instrumental rather than intrinsic values. The reason for this taxonomical decision brings us to the second issue, which is whether intrinsic value is to be identified with *final* value, i.e., the kind of value that pertains to that which is valuable irrespective of any considerations about conduciveness whatsoever. By way of illustration, consider the claim that true belief is non-instrumentally valuable in virtue of being a goal of inquiry, which, as argued above, implies that true belief is of non-instrumental *epistemic* value. Maintaining that true belief is valuable thus is perfectly compatible with also claiming that believing truly only is of non-instrumental epistemic value in so far as the beliefs at issue are of instrumental *non-epistemic* value (e.g., practical, prudential or moral), perhaps because they pertain to questions posed by some relevant set of inquirers.<sup>7</sup> In other words, true belief being of non-instrumental epistemic value does not imply that true belief is valuable thus independently of any considerations about conduciveness whatsoever.

Differently put, unlike bearers of final value, not all bearers of non-instrumental value are non-instrumentally valuable *simpliciter* or *without qualification* (such as “epistemic,” “moral,” or “practical”). Consequently, if intrinsic value is to be identified with final value, then the intrinsically valuable is *not* to be identified with the non-instrumentally valuable, as understood here. As already noted, being non-instrumentally valuable on *one* qualification does not imply being non-instrumentally valuable *without* qualification. As will be explored further in chapter 4, this is important to note not only because it shows that some non-instrumental values are not final values, but also because

---

of such value, but simply short-hand for saying that (at least some) instances of people believing truly or knowing things are epistemically valuable.

<sup>7</sup> See, e.g., Goldman (1999, pp. 94-96).

it suggest that it does not follow from believing truly being of non-instrumental epistemic value that *all* instances of true belief are of such value. Again, perhaps only those that pertain to questions posed by some relevant set of inquirers are valuable thus. But more on this in the next chapter.

For now, note that it not only does not follow from something being valuable on *one* qualification that it is valuable *without* qualification; it also does not follow from something being non-instrumentally valuable without qualification that it is non-instrumentally valuable on *every* qualification. In other words, something may be valuable *simpliciter* without being of non-instrumental epistemic value. By way of illustration, assume that undergoing an episode of pleasure is non-instrumentally valuable *simpliciter*. Even so, it does not follow that pleasure is of non-instrumental epistemic value, for the simple reason that pleasure is not an epistemic value in the first place. We will return to this point in §3.5 below.

To sum up, what follows from the above discussion is the following:

- (1) Some instances of true belief are of fundamental, non-instrumental epistemic value.

Note that it does not follow from (1) that all true beliefs that are non-instrumentally valuable are of *equal* non-instrumental value; some truths might very well be of greater non-instrumental value than others. Furthermore, that some true beliefs are of fundamental, non-instrumental epistemic value, clearly, does not imply that they are



*unique* in this respect—and this is a problem, in so far as we want to get from (1) to EVM. To do so, we need to also make plausible the following:

- (2) Nothing is of non-instrumental epistemic value, unless it is so in virtue of the non-instrumental epistemic value of believing truly.

If some instances of true belief are of fundamental, non-instrumental epistemic value, and nothing is of non-instrumental epistemic value unless it is so in virtue of the non-instrumental epistemic value of believing truly, then believing truly is unique in possessing fundamental, non-instrumental epistemic value—which, of course, is exactly what EVM states.

While (1) was motivated in the previous chapter with reference to how true belief is at least *one* goal of inquiry, defending (2) from potential objections will take further axiological legwork. To be exact, it will take the remainder of this chapter. It should be noted right at the outset, however, that there is a presumption in favor of conjoining (1) with (2), rather than conjoining (1) with a claim leaving room for a multitude of fundamental, non-instrumental, epistemic values. The presumption can be motivated with reference to a principle of axiological parsimony with respect to the number of kinds of bearers of fundamental value that is postulated. This principle is, in turn, an instance of the more general thesis that we should prefer ontologies with fewer rather than more existential commitments, *ceteris paribus*.<sup>8</sup> More specifically, the presumption is in favor

---

<sup>8</sup> Invoking this principle does not commit us to taking simplicity to be of fundamental, non-instrumental epistemic value, as opposed to of instrumental epistemic value (e.g., pursuing simplicity in theorizing is conducive to forming true belief), or non-epistemic

of explaining all judgments regarding absolute as well as comparative epistemic value by assuming one as opposed to several bearers of fundamental value. A pluralist might very well accept that there is such a presumption. What she would argue, however, is that this presumption is defeated by evidence to the effect that we cannot account fully for the domain of epistemic value in monistic terms, which is exactly the kind of charge that we will be concerned with below. Still, the presumption provides us with reason to prefer conjoining (1) with (2) in the *absence* of any convincing evidence for such an inability, and it is the burden of the remaining sections to show that there is no such evidence.

What constitutes relevant evidence here? (2) is equivalent to the universal claim that, for every  $x$ , if  $x$  is of non-instrumental epistemic value, then it is so in virtue of the value of believing truly. Given the obvious difficulty of defending a universal claim, I will restrict my attention to what, judging by recent work in epistemological axiology, are the most prominent *counterexamples* to (2). The counterexamples in question are formulated in terms of three candidates for fundamental, non-instrumental epistemic values, i.e., *knowing*, *being justified*, and *understanding*, to be discussed in §§3.2-3, 3.4, and 3.5, respectively. As we shall see, these candidates are typically taken to be backed up by strong and widespread axiological intuitions. This suggests that, if *any* candidate constitutes a challenge to (2), it will be found among these three. Consequently, if it can be shown that *none* of them poses such a challenge, we may conclude that there is no convincing evidence against (2), and that the presumption in favor of EVM still stands.

---

value (e.g., simplicity brings tractability, which is pragmatically valuable). However, see Sober (2001) for a skeptical take on the possibility of understanding the value of simplicity in terms of other values.

### 3.2. On the (Alleged) Surplus Value of Knowledge

What reason do we have for thinking that knowing is non-instrumentally valuable?<sup>9</sup> One reason is that knowledge is factive.<sup>10</sup> Given that some instance of true belief are non-instrumentally valuable, and knowledge implies truth, some instances of knowing may, thereby, be of derived non-instrumental value in virtue of being factive. The problem, however, is that many philosophers have wanted to make a claim not merely about knowing being valuable but about it being *more* valuable than believing truly. For example, Ward Jones (1997) writes that “I want all of my beliefs to be true—otherwise I would not believe them—but I would also rather they be known beliefs than mere true beliefs” (p. 424). And Jones is by no means alone. Duncan Pritchard starts out an overview of the literature on recent discussions regarding the value of knowledge by stating what he, correctly, takes to be an assumption shared by virtually everyone involved in the relevant discussions, namely that “we clearly do value knowledge more than mere true belief.” (p. 86). Moreover, many have argued that this idea can be found already in Plato’s *Meno*. Jonathan Kvanvig writes:

[...] part of the challenge of explaining the value of knowledge is explaining how it has more value than other things, one of these other things being true opinion—as Meno claims after acquiescing to Socrates’ point that true belief is every bit as useful as knowledge. “In that case, I wonder why knowledge should be so much more prized than right opinion” (97c-d). Meno expresses here a common presupposition about knowledge, *one that is widely, if not universally shared*. Given this presupposition, an account of the value of knowledge must explain more than how knowledge is valuable. It must also explain why the value of

---

<sup>9</sup> Here and henceforth, any unqualified statement about value concerns *epistemic* value.

<sup>10</sup> On one understanding of what knowledge is, knowledge just *is* true belief. We will not be concerned with this *weak* notion of knowledge here, but see Goldman and Olsson (2009) as well as Sartwell (1992) for two relevant discussions.

knowledge is superior to the value of true opinion (Kvanvig 2003, pp. 3-4; emphasis added).

We will return to the issue of what exactly Plato seems to be saying about the value of knowledge in the *Meno* in the next section. As far as contemporary discussions are concerned, the particular interpretation of aforementioned claims about the surplus value of knowledge that is relevant for present purposes is the following:

(K) Knowing is more *epistemically* valuable than merely believing truly.

After all, there might be many respects in which knowing is superior to mere true belief, including moral or practical. If it can be shown that knowing is valuable thus, that would be interesting and important.<sup>11</sup> But unless the surplus value in question is of a specifically *epistemic* kind, it would not speak to EVM, which is a thesis exclusively about epistemic value.

Could someone committed to EVM account for (K)? For one thing, (K) cannot be accounted for in terms of the non-instrumental value that derives from the factivity of knowing. Moreover, postulating an additional fundamental epistemic value, distinct from that of true belief, would clearly not be an option for a defender of EVM. Another option is, therefore, to argue that the non-factive component of knowledge contributes an additional *instrumental* value, since instantiating that component is conducive to believing truly. However, this strategy raises a problem that is typically directed against a particular account of what constitutes knowledge, namely reliabilism, but applies to any

---

<sup>11</sup> See, e.g., Olsson (2007) on the practical surplus value of knowledge.

axiological account that defines the non-factive component of knowledge in terms of truth-conducivity. Richard Swinburne sums up the central worry as follows:

Now clearly it is a good thing that our beliefs satisfy the reliabilist requirement, for the fact that they do means that [...] they will probably be true. But, if a given belief of mine is true, I cannot see that it is any more worth having for satisfying the reliabilist requirement. So long as the belief is true, the fact that the process which produced it usually produces true belief does not seem to make that belief any more worth having (Swinburne 1999, p. 58).

In other words, the problem is that the presence of truth seems to “swamp” the epistemic value of truth-conducivity, to use Kvanvig’s (2003) term. This does not mean that a defender of EVM cannot account for knowledge being epistemically valuable. It does, however, throw doubt upon her ability to account for (K).

Reliabilists have resisted this conclusion. For example, Alvin Goldman (Goldman and Olsson 2009) argues that (K) can be accounted for in terms of a pattern of inheritance by which token processes inherit value from states of true belief in virtue of being instances of process types that reliably produce such states. Moreover, he argues, the value inherited is *autonomous* in that it “isn’t dependent, on a case-by-case basis, on the value of resultant true beliefs” (p. 31) and may, as such, be added to the value that the relevant state of knowing has in virtue of its factivity. Presently, we will not be concerned with whether this is a successful solution to the swamping problem. Instead, it will be argued that any monistic attempt to account for (K) grants the defender of the swamping argument too much, since we lack any strong, pre-theoretic reasons for accepting (K) in the first place.<sup>12</sup> If we lack such reasons, there is no longer any pressure to surrender

---

<sup>12</sup> (K) has received some scrutiny in recent contributions to the debate, but not in a manner amounting to a rejection. Baehr (2009) raises doubt about the extent to which the

EVM for the purpose of accounting for (K) with reference to some non-instrumental value distinct from truth—or so it will be argued. Let us first spell out the case for the conditional, and then turn to the case for the antecedent.

Over the last couple of years, several analyses of knowledge have been motivated on the ground that they imply and, thereby, can account for (K).<sup>13</sup> Focusing on *pre-theoretic* reasons amounts to excluding reasons flowing from such implications. In fact, disqualifying theoretical reasons is in the interest of both parties of the debate, since the dialectical deck can be stacked either way. In particular, allowing for theoretical reasons to motivate or discredit (K) opens up for a very simple reply by the monist: Taken together, reliabilism and aforementioned swamping considerations imply that (K) does not hold; consequently, we have reason to believe that it does not. No one has pursued this strategy and the reason is simple: Those involved in the debate assume that (K) amounts to a constraint on *any* theory of knowledge, and acknowledge that reasons flowing from particular theories of knowledge carry no cross-theoretical force. The present focus on pre-theoretic reasons stays true to this dialectic.

That said, making the case that we lack any pre-theoretic reasons for accepting (K) is made complicated by the fact that it is hard to find any arguments for (as opposed to statements of) (K) in the literature. One relevant discussion can be found in one of the

---

intuition typically reported in the form of (K) really is fully general, in the sense of concerning *all* instances of knowledge, but seems to accept (K) as a rough generalization. Greco (2009) questions the idea that there is widespread intuitive support for the further idea that knowledge is more valuable than any subset of its constituents, while accepting (K). As will be discussed below, Kvanvig (2003) denies that knowledge is more valuable than any subset of its constituents, but accepts that (K) holds in virtue of the surplus value of justified true belief over mere true belief. Pritchard (forthcoming) initially rejects (K), but re-establishes it (at least in part) with reference to the value of understanding, arguing that knowledge typically goes hand-in-hand with understanding.

<sup>13</sup> See Pritchard (2007) for an overview.

earliest contributions to the contemporary debate regarding the value of knowledge, due to Jones (1997). According to Jones, one possible motivation for (K) is that it constitutes the best explanation of why so many philosophers have concerned themselves with analyzing knowledge. He writes:

Surely it is the high value we place on knowledge which has motivated the extraordinary volume of philosophical work which has gone into setting out the conditions for knowledge as opposed to true belief (Jones 1997, pp. 423-424).

The problem with this line of reasoning can be brought out by considering the fact that what any particular philosopher shows great interest in is a function of a whole host of factors, such as (a) what her colleagues happen to be showing great interest in, which influences the extent to which she will be successful in publishing her work or landing an academic position that will enable her to remain in the field; (b) what her supervisors happen to show great interest in, which influences the extent to which she will receive sufficient professional attention to develop necessary academic skills; and (c) what her philosophical predecessors happened to take great interest in, which influences the extent to which she will consider becoming a philosopher in the first place. This is not to suggest that philosophical inquiry is in no way sensitive to axiological facts, but merely to point out that establishing facts about comparative interests is not *sufficient* for establishing facts about comparative value, since what philosophers take great interest in is partly a function of a great many *sociological* facts that cannot be assumed to track relevant axiological facts about what degrees of value happen to apply to what states.

A second motivation for (K) may be gleaned from the fact that knowledge tends to be highly regarded in common discourse. Jones again:

Even more telling than philosophical infatuation is our unwillingness to endorse or applaud those who succeed in guessing. What is more, those who repeatedly guess tend to gain our disapproval, even though they may succeed at guessing. At the very least, we lose trust in their reliability as informants. Even more than valuing others' knowledge, we value ourselves as knowers. For most (if not all) of my factual beliefs, I take it to be important that they are knowledge and not lucky guesses. I want all of my beliefs to be true—otherwise I would not believe them—but I would also rather they be known beliefs than mere true beliefs (Jones 1997, p. 424).

To see the problem with this line of reasoning, we need to distinguish between two targets of evaluation, i.e., doxastic *states* and belief-forming *processes*. With this distinction in mind, consider, first, the fact that we do not applaud guessing. Construed as a process, this could be explained with reference to the unreliability of guessing, and how the process, hence, is of instrumental disvalue. But what about the doxastic states ensuing from guessing? A lucky guess amounts to something of non-instrumental value, i.e., a true belief, which, furthermore, explains why we value lucky over unlucky guesses, *ceteris paribus*. Notice, however, that it makes little sense to talk about a *process* of lucky guessing, as opposed to one of guessing *simpliciter*. After all, what would such a process involve? Surely, it would not involve consistently getting things right; in that case, it would no longer be plausible to refer to the process as involving guessing, as opposed to some sub-personal sensitivity that (save for its reliability) might resemble guessing in not being preceded by any deliberation. Indeed, as with any form of guessing, the process would have to be one that is generally unreliable and, as such, one that we have little reason to value.

Next, turn to the claim that we value ourselves as knowers. Construed as a claim about how we value knowing things, this can be accounted for with reference to the



factivity of knowing. The same holds if we construe the claim as being about how we value engaging in processes that involve us coming to know things. Consider, next, the comparative claim that we value knowledge more than we value lucky guesses. On a state reading, it is not clear that this claim is true. Imagine two scenarios. In one scenario, you make a guess and form a true belief as a result. In the second scenario, you make a reliable inference and, as a result, form a true belief that, moreover, amounts to knowledge. Now, consider the comparative value *not* of your means of forming belief but merely of the two resulting states. Is one really more valuable from an epistemic point of view than the other? It is not clear that it is, given that they both involve true belief.<sup>14</sup> In so far as we intuit a difference here, it is likely that the intuition involved tracks what *does* seem clear, namely that the *processes* involved in arriving at the respective states are not equally valuable. The procedure utilized in the first scenario is of instrumental disvalue in virtue of being unreliable, while that of the former is of instrumental value in virtue of being reliable.

This brings us to the final claim of Jones', i.e., that "I would also rather [my beliefs] be known beliefs than mere true belief." This is simply a restatement of (K), i.e., the very claim for which we are seeking a motivation. If it were the case that the preceding claims could only be accounted for by assuming (K), then we could have considered the statement a hypothesis, needed to account for the truth of the relevant axiological claims. But what we have just seen is that it is perfectly possible to account for those claims within a monistic framework in the manner just outlined. Consequently, Jones' assertion of (K) amounts not to an explanatory hypothesis necessary for explaining

---

<sup>14</sup> Here, I'm siding with Kvanvig (2003, p. 148).

the observations preceding it, but simply to the statement of a preference. This is certainly not to call into question the sincerity of Jones' remark; it is simply to ask whether a single, reported preference is sufficient to warrant acceptance of (K).

This brings us to the third potential motivation for (K), gleaned from the observation that Jones' preference is fairly *widespread*, as noted already above. However, it is questionable whether this provides any *pre-theoretic* reason for accepting (K) which, as we saw above, is the kind of reason that needs to be provided. For one thing, we would need evidence to the effect that the intuition in question is not merely a function of previous theoretical commitments. This is not to defend some general form of skepticism regarding the evidentiary force of intuitions in philosophical inquiry. It is merely to suggest that intuitions have to be bolstered by supporting considerations (as opposed to simply being taken at face value) exactly to the extent that they (*a*) pertain to subtle philosophical matters and (*b*) are leveraged against reputable philosophical theories. Since matters of comparative epistemic value are subtle, and EVM is a reputable theory (which is something that has to be granted even by those who take it to be false), a question arises: Is it the intuition that knowledge is more valuable than mere true belief that provides the impetus for coming up with a theory that satisfies this intuition, or is it the having of a theory on which knowledge is more valuable than mere true belief that provides the impetus for the intuition? The absence of any arguments ruling out the latter should give us pause before accepting (K) as a pre-theoretically motivated constraint on investigations into the value of knowledge.

### 3.3. Meno and the Monist

There is one final place where the defender of (K) might seek pre-theoretic support for the idea that knowing is more epistemically valuable than merely believing truly, even in the absence of contemporary pre-theoretic support. That place is Plato's *Meno*. In fact, Socrates' observations in the *Meno* about the surplus value of knowledge over mere true belief has become such a common point of reference for defenders of (K) that the problem of explaining why (K) holds often is referred to simply as *the Meno problem*.<sup>15</sup>

That said, this taxonomy is misleading on two points. First, as pointed out by Dennis Whitcomb (forthcoming), it is misleading to talk about *the Meno problem*. In fact, Whitcomb argues that, depending on exactly how the relevant surplus claim is spelled out, there are at least 540 different problems that one can mean to be addressing. This, of course, is compatible with it being the case that Plato and those that turn to him in support of (K) still have the very same problem in mind. But this brings us to the second point: As far as we are concerned with (K), it is even misleading to talk about the *Meno problem*. More specifically, it will be argued that, to the extent that there is a surplus claim in the *Meno*, it is a claim that differs from (K) with respect to both the *scope* of the relevant claim and the *kind* of value that is at issue. Consequently, the defender of (K) has little to gain from turning to the *Meno* for support.<sup>16</sup>

---

<sup>15</sup> Here and henceforth, any reference to Socrates refers to the Socrates of Plato's dialogues, not to the historical Socrates.

<sup>16</sup> It will be assumed presently that the term *episteme* is best translated as "knowledge." This translation has supporters in the literature (e.g., Fine 2004), but is not altogether uncontroversial (see, e.g., Burnyeat 1980). Notice, however, that if *episteme* should *not* be translated as "knowledge," that is just yet another reason not to invoke the *Meno* in support of (K).

Consider, first, the scope of the surplus claim that figures in the *Meno*. Towards the end of the dialogue, Socrates suggests that knowledge is no more *useful* than mere true belief (96c), and illustrates his point with reference to how someone knowing the way to Larissa will be no worse of a guide to others with respect to getting them there than someone who merely has a true belief about the way to Larissa (97a). This, famously, makes Meno wonder “why knowledge is prized far more highly than right opinion, and why they are different” (97c-d).<sup>17</sup> In response, Socrates invokes the idea that knowledge is true belief tied down by reason:

[...] true opinions, as long as they remain, are a fine thing and all they do is good, but they are not willing to remain long, and they escape from a man’s mind, so that they are not worth much until one ties them down by (giving) an account of the reason why. And that, Meno, my friend, is recollection, as we previously agreed. After they are tied down, in the first place they become knowledge, and then they remain in place. That is why knowledge is prized higher than correct opinion, and knowledge differs from correct opinion in being tied down (97e-98a).

This, of course, is the very passage that has led many contemporary epistemologists to suggest that Plato subscribed to something along the lines of (K) in the *Meno*. However, interpreting the passage thus is problematic. Notice that Plato identifies the reasoning<sup>18</sup> relevant to knowing with *recollection*. Earlier in the dialogue, the process of recollection is invoked in a solution to the so-called paradox of inquiry, introduced by Meno and reformulated by Socrates in terms of a dilemma. According to the first horn, the inquirer “cannot search for what he knows—since he knows it, there is no need to search” (80e).

---

<sup>17</sup> All translation of Plato are from Cooper (1997).

<sup>18</sup> As indicated by the parenthetical remark in the translation “(giving) an account of the reason why,” the relevant Greek phrase “*aitias logismos*” is generally thought to refer not to the product but to the *process* of reasoning (*logismos*) to the cause or explanation (*aitia*). See Fine (2004, p 58) for a discussion.

On the second horn, the inquirer also cannot search “for what he does not know, for he does not know what to look for” (*ibid.*).

Socrates’ solution to the paradox, involving the famous interrogation of the slave boy about how to double the area of a square (82b-85c), is that “the man who does not know has within himself true opinions about the things that he does not know” (85c) and that, “if he were repeatedly asked these same questions in various ways, you know that in the end his knowledge about these things would be as accurate as anyone’s” (85c-d). In other words, the reason we can engage in philosophical inquiry is that we have tacit yet true belief.<sup>19</sup> All it takes is an elenchus. Through a persistent, dialectical inquiry in the form of question and answer adversary argument, we may recollect what we used to know when disembodied, but that has since been demoted from explicit knowledge to tacit but nevertheless true belief.

That said, the claim that the theory of recollection is invoked by Plato to solve the problem of inquiry is not altogether uncontroversial. For example, Dominic Scott (2006) argues that the theory is instead invoked in a psychological strategy of carrot and stick, attempting to both exploit Meno’s weakness for the exotic through allusions to ancient myths in a context where he is getting tired of Socrates’ constant refutations, and arouse a sense of shame in Meno by suggesting that he is weak for wanting to leave the

---

<sup>19</sup> Some commentators (e.g., Taylor 2008 and Scott 2006) would prefer to talk about tacit *knowledge*, and point to how Socrates (at 85d) talks about “the knowledge he [i.e., the slave boy] now possesses”. Other commentators (e.g., Fine 2003) suggest that Socrates here is referring to an envisaged future time, at which the slave boy has undergone a successful elenchus, and point to how Socrates denies (at 85c) that the slave boy knows anything at the end of the interrogation. Presently, I’m working with the latter interpretation. Note, however, that the arguments to be made can be made just as easily on the former interpretation as on the latter.

discussion.<sup>20</sup> To the extent that the theory of recollection is introduced to solve any philosophical problem, Scott argues, it is introduced to address a challenge that (Scott maintains) is ignored in Socrates' reconstruction of Meno's challenge, when the latter asks, "if you really stumble upon it [i.e., the definiens], how will you know that this is the thing you didn't know before?" (80d). Scott refers to this question as posing a *problem of discovery*.

There is no reason to doubt that Plato introduced the theory of recollection to play a subtle rhetorical role in relation to Meno. It is less clear, however, whether it was not *also* meant to solve the problem of inquiry. After all, what Scott refers to as the problem of discovery seems a mere special case of the problem posed by Socrates' second horn, i.e., that the inquirer cannot search "for what he does not know, for he does not know what to look for" (80e).<sup>21</sup> On this point, notice that Meno does not resist Socrates' reconstruction (81a). And why should he? Someone who does not know what he is looking for has several problems as far as inquiry is concerned, *one* of which is that he will not be able to tell if he has stumbled upon the target of investigation—unless his inquiry is guided by tacit true belief, as postulated by the theory of recollection. That is why inquiry is *possible* and the inquirer avoids the second horn. Moreover, given that the beliefs guiding him do not constitute knowledge until tied down through the relevant process of reasoning (i.e., through recollection), inquiry is still *worthwhile* and the inquirer avoids the first horn.

---

<sup>20</sup> Hence, Socrates suggests that "we would be better, more manly and less lazy if we believed that we ought to inquire into what we do not know, than if we believed that we cannot discover what we do not know and so have no duty to inquire" (86b-c).

<sup>21</sup> On this point, notice that Meno does not resist Socrates' reconstruction (81a).

All of this is relevant presently for the following reason: If (a) the kind of knowledge that Plato has in mind when having Meno suggest that knowledge is prized far more highly than mere true opinion is the kind of knowledge that results from true belief being tied through a process of reasoning; (b) to undergo the relevant process of reasoning is, as Socrates suggests, to recollect; and (c) recollection is introduced for the purpose of explaining why philosophical inquiry is possible, then the kind of knowledge that Plato is concerned with in the *Meno* is not just any kind of knowledge; it is the particular kind of knowledge had by those who have gained successful insight into the nature of things. In short, it is *philosophical* knowledge, not knowledge in general.

This is *not* to suggest that Plato was committed to a theory of Forms already in the *Meno*. It suffices to note that the one theme that is recurrent throughout the dialogues is an inquiry into *essences*, or that by which the phenomena under investigation are what they are.<sup>22</sup> Given the epistemology of recollection introduced in the *Meno*, interpreting that by which virtues “have one and the same form which makes them virtues” (72c) in terms of proto-Forms is tempting but not necessary for present purposes. The only thing maintained here is that, as far as the *Meno* is concerned, claims about the surplus value of knowledge apply to the kind of knowledge relevant to recollection, and the kind of knowledge relevant thus is philosophical knowledge, as in knowledge of essences,

---

<sup>22</sup> Hence, when Socrates asks in the *Meno* “If I do not know what something is, how could I know what qualities it possesses?” (71b), he not only seems to appeal to an implicit distinction between essential and non-essential properties, but also follows up the request for an account of the latter (at 72b) by asking for the nature of something, using the same word (*ousia*) as he uses in the *Euthyphro* (at 11a-b) when asking for the essence of piety. The same distinction between the essential and the non-essential appears in the *Laches* (189e-190c), *Protagoras* (312c and 360e-361d), and *Gorgias* (463c).

whether or not such essences are to be spelled out in terms of Plato's later ontology of Forms, or in terms of his earlier ontology in terms of essences more generally.

Moreover, maintaining that the *Meno* is exclusively concerned with philosophical knowledge is also not to suggest that Plato necessarily took philosophical knowledge to be the *only* kind of knowledge there is.<sup>23</sup> If he did, then no qualification would be needed, since there would no longer be a distinction between philosophical knowledge and knowledge in general. Be that as it may; the only things necessary to make the present point is that (a) contemporary epistemologists looking to the *Meno* in support of a claim about the surplus value of knowledge do *not* restrict their claims to philosophical knowledge, despite the fact that (b) the kind of knowledge at issue in the *Meno*'s discussions about the value of knowledge is philosophical and *only* philosophical knowledge, whether or not Plato at any point took that to be the only kind of knowledge there is.

One potential counterexample to (b) is the fact that the distinction between true belief and knowledge is introduced in the *Meno* with reference to knowing the way and guiding others to Larissa (97a). Taken literally, this example suggests that Plato's concerns in the *Meno* were not exclusively with philosophical knowledge. However, consider that the topic of the dialogue, i.e., *arete*, which is typically translated as 'virtue,' is an inherently *political* concept as far as the *Meno* is concerned.<sup>24</sup> In light of this, it is not clear that we should understand the Larissa example as anything but a mere metaphor for the kind of leading or guidance that Socrates ultimately seems to be concerned with,

---

<sup>23</sup> There is some indication to the effect that Plato took knowledge to be restricted to knowledge of forms in the *Republic* (see particularly the end of book V and books VI-VII). See Taylor (2008, pp. 178-9) for a discussion.

<sup>24</sup> See Scott (2006, p. 14) for a discussion.



i.e., the kind that Themistocles *et al.* are engaging in when leading their cities (99b). After all, Larissa only figures twice in the dialogue; first, as a place inhabited by people admired for their wisdom (70b), and, then, in aforementioned example, illustrating the absence of a distinction between knowledge and mere true belief as far as usefulness is concerned. At no point in the dialogue does Socrates concern himself with the kind of knowledge that literal, geographical location and guidance would have to pertain to, i.e., empirical knowledge. Consequently, it is far from clear that Socrates would say that we may literally *know* the way to Larissa, in the way that we (if genuinely virtuous) may know how to lead a city, which is the kind of knowledge occupying the discussion at large.<sup>25</sup>

As such, we may conclude that, to the extent that Plato subscribes to anything like (K) in the *Meno*, he is, at most, committed to the following:

(K\*) Having *philosophical* knowledge is more epistemically valuable than is merely believing truly.

Understood thus, the problem as far as invoking the *Meno* in contemporary debates about the value of knowledge is concerned, is that those invoking (K) in an argument against EVM, as already noted, are concerned with knowledge *in general*. As such, there is a clear difference in *scope* between the surplus claim found in the *Meno* and that defended

---

<sup>25</sup> A similar point can be made about the passage at the very beginning of the dialogue where Socrates talks about knowing who Meno is (71b), which, as Scott (2006, p. 21) points out, “is best treated as a pedagogical device to give Meno an intuitive hold on the idea of one question (‘what is x?’) having priority over another (‘what is x like?’).”

by recent epistemologists. Consequently, it is not clear that the passages typically quoted from the *Meno* provide great support—pre-theoretic or otherwise—for (K).

In fact, further consideration of what Plato has Socrates say about the value of knowledge suggests that the former's take on the value of knowledge not only differs from (K) with respect to scope, but also with respect to the *kind* of value involved. According to Socrates, (philosophical) knowledge is more valuable than mere true belief for being tied down through a process of reasoning about the relevant essences in virtue of which the knowledge at issue, unlike the untied work of Daedalus, will “remain in place” (98a). In other words, philosophical knowledge is more valuable than mere true belief (philosophical or otherwise) on account of the former's *stability*. This begs a further question: Why is stability valuable? Surely, it cannot be because it makes philosophical knowledge more fit for use in action, since Socrates explicitly denies that there is any difference between knowledge and mere true belief on this score (98c).

A more plausible interpretation stays true to the central theme of the *Meno*, i.e., the relation between possessing philosophical knowledge and being able to *teach* virtue. The connection between teachability and the stability of the former is not immediately obvious from Socrates' brief remarks in the *Meno*, but we may begin to see what is going on if we follow Gail Fine (2004) in looking beyond the *Meno* toward the *Republic*. According to Fine, the kind of stability at issue in the *Meno* should be understood in terms of what is said in the *Republic* (540b-c) about how someone who knows (at least in the limiting case of someone grasping the Good itself), unlike someone who merely

believes truly, is invulnerable to refutation.<sup>26</sup> The same goes, of course, for someone who just happens to be extremely dogmatic, which suggests that what is valuable about the relevant kind of stability is not invulnerability *per se*, but rather the particular kind of invulnerability that arises from judging things “in accordance with being” (543c). Differently put, philosophical knowledge is not valuable because it is tied down, but because of what it is tied down *to*, i.e., an account of the essence of things.

Why is it valuable for knowledge to be tied down to such accounts? As Socrates proposes in the *Meno*, having one’s beliefs be tied down thus is to have philosophical knowledge, and to have philosophical knowledge is, as Socrates goes on to suggest, to be able to teach virtue and, in effect, “make another into a statesman” (100a). In other words, philosophical knowledge is (at the very least) valuable in that it enables one to *multiply* virtue, which raises a further question: Why is it valuable to be virtuous? Throughout the *Meno*’s discussion about the value of virtue (87d-89a), the only value it is explicitly said to have is that which it has on account of enabling one to use one’s assets correctly and, in effect, produce happiness (*eudaimonia*, at 88c). In modern parlance, virtue is, thereby, *instrumentally* valuable, in virtue of its conduciveness to happiness.

In other words, as far as the *Meno* is concerned, both philosophical knowledge and philosophical true belief is instrumentally valuable, in that it enables one to use one’s assets correctly and, thereby, produce happiness. However, having philosophical knowledge is *more* valuable than having true belief (philosophical or otherwise) because someone endowed with the former not only will be able to use her assets correctly, but

---

<sup>26</sup> See also *Euthyphro*, where Euthyphro complains that “whatever proposition we put forward [for cross-examination] goes around and refuses to stay put where we establish it” (11b), which prompts Socrates to make an allusion to Daedalus (11b-e).

also be capable of instilling the same capacities in others and, thereby, bring about happiness on a much greater scale. More specifically, in so far as one has philosophical knowledge, one will in general be able to bring about a much greater amount of happiness, since the beneficial consequences of one's endowment extends beyond the (correct) use of one's own assets, to include also the aggregate benefits of the actions of one's students of virtue. Consequently, philosophical knowledge is prized for more highly (as Meno puts it) than mere philosophical true belief simply in virtue of how bringing about more good is better than bringing about less good.

Aforementioned line of reasoning makes sense of Socrates' approval of Meno's suggestion that philosophical knowledge tends to be prized far more highly than mere philosophical true belief, as well as his explanation in terms of stability and recollection. It does not go to show, however, that every instance of philosophical knowledge is more valuable than every instance of mere philosophical true belief. After all, we can surely imagine a statesman of mere true (philosophical) belief bringing about an immense amount of good on account of ruling a massive city correctly, while another statesman, albeit endowed with knowledge, is unable to find neither subjects nor students on account of geographical isolation or other external circumstances. However, as already noted, what Socrates sets out to show is merely that the former is prized far more highly than the latter. To make that claim plausible, it is sufficient to show that philosophical knowledge *in general* gives rise to more happiness than mere true belief, philosophical or otherwise.

Notice, however, that the above interpretation of Plato's take on the value of knowledge in the *Meno*, interestingly, gives rise to a kind of swamping problem, pertaining not to conduciveness to truth but to happiness. More specifically, if the value

of virtue is exhausted by the instrumental value it has because it is conducive to happiness, does the presence of happiness swamp the value of virtue? It is not clear what Plato would say here at the time of the *Meno*. It is interesting to note, however, that Plato in the *Republic* sets out to defend the claim that justice is not only good as an instrument but also as an end, in being both an instrument to happiness and an integral part of the good life. However, no analogous claim is made about the value of virtue in the *Meno*, which merely talks about the instrumental value of being virtuous. As such, I will leave it open whether such a swamping problem would worry Plato at the time of the *Meno*.

This leaves us with one final question about Plato's take on the value of knowledge in the *Meno*: What *kind* of value applies to philosophical knowledge? If philosophical knowledge is valuable because it enables the knower to multiply virtue, and virtue is valuable because it enables one to use one's assets correctly and, in effect, produce happiness, it seems fair to say that, whatever value attaches to knowledge on account of being generally conducive to happiness thus, it is *not* an epistemic kind of value. In other words, in so far as Plato subscribed to anything like (K) at the time of the *Meno*, the most plausible interpretation is the following:

(K\*\*) Having philosophical knowledge is more *non-epistemically* valuable than merely believing truly.

Setting aside the issue that (K\*\*), like (K\*), only pertains to *philosophical* knowledge, as opposed to knowledge in general, notice that (K\*\*) is compatible with EVM, which is a thesis exclusively about epistemic value. More generally, given that it does not follow

from something being valuable on *one* qualification (e.g., *morally* valuable, *prudentially* valuable, or what have you) that it is valuable on every other qualification, (K\*\*) is compatible with EVM no matter how the relevant non-epistemic value is spelled out. In fact, given that it does not follow from something being valuable *without* qualification that it is valuable on *every* qualification, even a construal of (K\*\*) in terms of knowledge simply being more valuable *simpliciter* than mere true belief is compatible with EVM. In conclusion, it is far from clear that contemporary critics of EVM have much to gain in terms of support from Plato's *Meno*.

#### 3.4. Justification and Non-instrumental Value

The previous sections suggested that there is scant pre-theoretic support for (K) in the contemporary literature, and far less to gain for the defender of (K) in turning to the *Meno* than what seems to be assumed in the literature. As such, there are no strong reasons to account for (K), let alone to account for it in terms of knowledge being of fundamental value. However, consider a point made recently by Jason Baehr (2009), who argues that it is implausible to believe that, “when we think of knowledge as more valuable than true belief at the relevant intuitive level, we are thinking of it *purely* in the abstract, without any (even implicit) reference to any of the features *in virtue of which* it is apparently more valuable” (p. 51). This highlights a possibility ignored in the previous sections, namely that the intuition behind (K) does not concern the surplus value of knowledge directly, but rather the fundamental value of some *component* of knowledge,

that, when combined with the value of the factivity of knowledge, amounts to knowledge being more valuable than mere true belief.

What are the candidates for such a component? Clearly, truth is not one of them, nor is mere belief. This leaves us with two partially overlapping candidates, namely justification and warrant. Here, warrant is understood as whatever factor turns true belief into knowledge (Plantinga 1993), and justification as warrant minus whatever condition needs to be added to a JTB account in order for it not to have Gettier cases come out as cases of knowledge. Presently, the focus will be on justification, the reason being that it is unclear whether an account of warrant will tell us anything interesting about value. In a passage commenting on Peter Klein's (1981) attempt to avoid the Gettier problem by way of a distinction between different kinds of defeaters, Kvanvig writes:

It is hard to see this distinction as anything more than gerrymandering needed to prevent counterexamples to one's account of knowledge, and it is easy to side with Williamson in remarking 'Why should we care about *that*?' The distinction between these kinds of defeaters tracks no intuitive difference in value, leaving us with an account of the nature of knowledge incapable of helping to explain the value of knowledge (Kvanvig 2003, pp. 129-130).

According to Kvanvig, Klein's analysis is in no way unique in this respect. In fact,

[...] the better an approach to the Gettier problem is at carving cases of knowledge off from cases of non-knowledge, the more *ad hoc* and gerrymandered the proposal. The result is a condition which has no hope whatsoever of giving a decent answer to the question of what makes un-Gettiered justified true belief more valuable than justified true belief. Hence, the hope of defending the view that knowledge is more valuable than any proper subset of its parts is dim indeed (Kvanvig 2009, p. 103).

In other words, any account that includes a Gettier clause (which any account of warrant will) is bound to be *ad hoc* and gerrymandered, and any *ad hoc* and gerrymandered

analysis will fail to explain the surplus value of knowledge over justification. This suggests that the most promising place to look for an account of the surplus value of knowledge over true belief is not in the Gettier clause. Furthermore, since the Gettier clause is what differentiates warrant from justification, focusing on the latter rather than the former seems a more fruitful strategy.<sup>27</sup>

In fact, Kvanvig argues that there are at least two kinds of accounts of justification that may escape the swamping problem and, thereby, account for (K), namely certain (access) internalist and virtue epistemological theories. Further examples can be found. As we saw in the previous chapter, Richard Feldman (2002) denies that believing truly as opposed to being justified is of non-instrumental epistemic value. If Feldman is right, (K) holds simply in virtue of the fact that (merely) believing truly is of no non-instrumental epistemic value, while knowing is of such value in virtue of the non-instrumental value of being justified. Since I have already discussed Feldman's position at length, I will not discuss it further here. Moreover, it is important to remember that what we are after here are *pre-theoretic* reasons for at all seeking to account for (K). As stressed above, simply providing a theory that implies that (K) holds, without providing any reasons as for why theories *should* imply (K), would be in the interest of neither party of the debate.

With this in mind, return to the possibility called attention to above about the intuition captured by (K) being an intuition not about knowledge *per se*, but rather about

---

<sup>27</sup> This is *not* to suggest that an analysis being “ugly” yet correct implies that the analysandum cannot be valuable. As pointed out by DePaul (2009), we cannot assume that accuracy of analysis implies transparency with respect to value. The point relevant here is simply that consulting analyses that are bound to be “ugly” does not constitute a promising strategy, in so far as we are interested in questions about value, exactly *because* we cannot assume such transparency.



the fundamental value of some *component* of knowledge, that, when combined with the value of the factivity of knowledge, amounts to knowledge being more valuable than mere true belief. Then assume, if only for the sake of the argument, that there is, in fact, a widespread, pre-theoretic intuition to the following effect (for some sufficiently pre-theoretic notions of ‘justification’ and ‘non-instrumental’):

(J) Being justified is non-instrumentally valuable.

A widespread intuition to this effect would constitute *prima facie* evidence for the claim that being justified is non-instrumentally valuable. Moreover, given that that justification is non-factive, the non-instrumental value in question cannot be derived from that of true belief, as in the case of knowledge and other factive states. Consequently, the most plausible response to the evidence would be to consider the relevant value fundamental. As such, the evidence in question would not only provide indirect support for (K) but also constitute a direct challenge to (2), i.e., the idea that nothing is of non-instrumental epistemic value unless it is so in virtue of the non-instrumental epistemic value of believing truly.

That said, the *prima facie* evidence in question might, like any *prima facie* evidence, be defeated. In fact, it will be argued momentarily that the evidence in question *is* defeated. More specifically, it will be argued that it cannot be ruled out that we would intuitively take being justified to be non-instrumentally valuable even if it was not, given what can reasonably be assumed about the central role that the inculcation and reinforcement of an appreciation for instrumentally valuable states and processes plays in

inquiry. As such, the argument to be provided discredits the reliability of the particular intuition in question, as far as the tracking of relevant axiological facts about the value of justification is concerned, and suggests that even a widespread intuition to the effect that (J) holds does not provide a reason for taking its content at face value.

First, however, it needs to be established that the relevant kind of axiological confusion is at all conceptually possible. After all, if what it *is* for something to be valuable simply is for it to be valued (or preferred) in the relevant way, the confusion in question is not possible. However, there are two problems with such a preferentialist account of epistemic value. First, it does not do a good job of accounting for instrumental value. Consider, for example, a particular research method. In so far as using that method is conducive to something that is valuable for the inquirer, such as true belief, using the method will be instrumentally valuable for the inquirer. However, it does not follow that such usage, thereby, is instrumentally *valued* by the subject—after all, she may not even be aware of the method, let alone of its virtues. Consequently, preferentialism is not a plausible theory for instrumental epistemic value.

In fact, the thesis of preferentialism remains questionable, even if restricted to non-instrumental value. The main problem is that such a thesis would either beg the question against a defender of (J), or run the risk of rendering the domain of epistemic value unrecognizable. To see why, consider that the thesis would be that all and only targets of non-instrumental preferences constitute non-instrumental values. Specified thus, there is no constraint whatsoever on what *kinds* of preferences would be relevant to epistemic as opposed to other kinds of value. In particular, there would be no way to disqualify completely arbitrary preferences, such as a non-instrumental preference for

beliefs formed on Wednesdays. Or rather, there *is* a possible constraint, to the effect that the preferences in question must be connected to considerations about truth. However, that would simply beg the question against those who want to take the intuition to the effect that (J) holds at face value and maintain that justification is valuable *independently* of considerations about truth.

Consequently, it should come as no surprise that preferentialism typically is identified with the significantly weaker thesis that non-instrumental value accrues *only* to (but not to *all*) targets of non-instrumental preferences.<sup>28</sup> Notice, however, that since such a preferentialism does not take non-instrumental preferences to be sufficient for non-instrumental value, it does not rule out the possibility that an inquirer values something non-instrumentally, even if that something is only instrumentally valuable.

But conceptual possibility is one thing; in order to speak to the question of whether we should take an intuition to the effect that (J) holds at face value, we need to say something about whether it is at all likely that the relevant kind of confusion—i.e., a confusion to the effect that we mistakenly have come to consider justification non-instrumentally valuable, despite it only being instrumentally valuable—might have taken place and is driving the intuition in question. Above, it was suggested that epistemic value is to be understood in terms of the goals of inquiry. As such, it seems reasonable to suppose that our intuitions about epistemic value will be sensitive to considerations about the manners in which such goals are pursued in the practice of inquiry. With this in mind, consider that we neither have voluntary influence over what we believe, nor an ability to determine directly whether what we believe is true. Indeed, even such an optimist about

---

<sup>28</sup> See, e.g., Rønnow-Rasmussen (2002).

introspection as Descartes took to developing rules for the direction of mind. Today, we may have our doubts about the particular rules that he settled on, but the general insight remains: we want truth, but in a “nonmagical world” like ours, to borrow a phrase from Bernard Williams (1978), we have no choice but to resort to pursuing *means* for increasing the chances that we will come to believe truly.

This predicament has implications for what kind of norms of inquiry can be fruitfully used and communicated in the process of epistemic education. The first implication is a negative one: Norms formulated exclusively in terms of true belief, e.g., “Believe what is true,” will not only have an air of complete triviality about them—“*Of course* we should believe what’s true,” the inquirer in training might say, and rightly so in light of above reflections on the goal of inquiry—but also be completely unhelpful from the point of view of teaching a budding inquirer how to conduct inquiry successfully. The second implication is a positive one: The kind of norms that may be fruitfully communicated in the process of epistemic education is exactly the kind that pertains to means to truth, rather than to truth itself. This is mirrored by the way in which inquiry (in its many manifestations) is typically taught. What is inculcated is not only a love of truth, but also—and perhaps even more persistently—a love for a variety of processes and activities of epistemically informed conduct, be it in the form of the sometimes quite abstract principles about evidence, reflection, and coherence often found in epistemological analyses, the variety of logical tools taught in philosophy classes, or the rigorous methods of scientific inquiry used in the natural sciences.

In fact, it seems reasonable to believe that the majority not only of the inquirer’s education but also of her subsequent inquiry will be concerned not with questions of truth

*per se*, but rather with the pursuit of means to truth. After all, the reason norms formulated exclusively in terms of true belief are of so little use to inquiry is not merely that they play a very limited role in the process of education, but also because they are largely unhelpful for the practice of inquiry in general. More specifically, for any question pondered by the inquirer, she either knows the answer, or she does not. If she knows the answer, any injunction to the effect that she should believe what is true will be *unnecessary*. If she does not know the answer, the same injunction will be *useless*. For this reason, the love for the means to truth that is inculcated throughout an inquirer's education is likely to be further reinforced throughout her practice, and it cannot be ruled out that she over time comes to value those states in themselves.

This hypothesized process of inculcation and reinforcement has a lot in common with what Goldman (Goldman and Olsson, 2009) recently has referred to as *value autonomization*, whereby processes that originally are regarded as merely instrumentally valuable later are “upgraded to the status of independent value, thereby accommodating the legitimacy of *adding* their value to that of true-belief outcomes” (p. 35). As we saw above, Goldman invokes this process in order to account for why we ascribe value in accordance with (K). However, in the preceding sections, doubts were raised as to whether we really have any pre-theoretic reasons to account for (K) in the first place. The purpose of invoking an analogous process here is, therefore, *not* to account for (K), but rather to *undercut* the claim that an intuition to the effect that (J) holds should be taken to provide indirect, pre-theoretic support for (K), by virtue of justification being a component of knowledge.

As such, the present argument is more directly related to John Stuart Mill's (1861) argument against virtue theoretic counterexamples to the principle of utility. Mill famously maintained about virtue that "[t]here was no original desire of it, or motive to it, save for its conduciveness to pleasure, and especially to protection from pain" (p. 38). However, due to that very conduciveness, we have come to not only associate the virtues and vices with the moral and the immoral, respectively, but over time also forgotten the yardstick against which the virtues originally earned their titles. Mill makes an analogy with the love of money, noting that, even if the worth of money is solely that of the things which it will buy, "money is in many cases desired in and for itself; the desire to possess it is often stronger than the desire to use it, and goes on increasing when all the desires which point to ends beyond it, to be compassed by it, are falling off" (p. 37). The possibility pondered presently is that something similar might have happened with our axiological conceptions of justification. As aforementioned reflections on preferentialism revealed, however, such a re-conceptualization does not *make* justification non-instrumentally valuable. Albeit valued in itself—and probably for good reason, given how inquiry is taught and practiced—the means associated with justification remain of mere instrumental value, in so far (and *only* in so far) as they enable the inquirer to form true belief.

Naturally, none of this goes to establish beyond all possible doubt that such axiological confusion has actually taken place. Still, the purpose of the present discussion is merely to suggest that such confusion cannot be ruled out, given what we can reasonably assume about the way in which inquiry is typically taught and practiced. This ties in with what was suggested above about (K), namely that intuitions that pertain to

subtle philosophical matters and are leveraged against reputable philosophical theories have to be bolstered by supporting considerations, as opposed to simply being taken at face value. It is in this context that the above line of reasoning should be understood. By providing a plausible story suggesting that any intuition to the effect that (J) might not track the relevant axiological facts, the argument puts a significant burden of proof on those who want to take (J) at face value. In the absence of a rival account of the source of the intuition in question that explains why it would track the relevant facts, we have no reason to endow intuitions to the effect that (J) holds with any great evidential weight, or take it to provide a convincing counterexample to (2).

As above, it bears stressing that this is not a case for general skepticism regarding the evidentiary force of intuitions, but merely a suggestion to the effect that a particular intuition cannot be taken at face value, given what we can reasonably assume about its potential insensitivity to the relevant facts. Granted, a structurally similar argument could, in principle, be given for other candidates for non-instrumental value as well. However, two things should be noted. First, each such additional argument would have to be considered on its own merit, with respect to how plausible a story it can provide about how the intuitions in question fail to track the relevant axiological facts. In other words, while the above argument may provide a *structure* for such arguments, it certainly does not provide their *conclusions*. Second, interestingly, it is not clear that such an argument could plausibly be given with respect to the one value on which the present investigation rests, i.e., the non-instrumental value of true belief. After all, such an argument would have to conclude that true belief is not even *a* goal of inquiry—a claim that could be interpreted in either of two ways. On the one hand, it could be understood as a claim to

the effect that there is a multitude of incompatible yet equally legitimate notions of inquiry, only one of which is defined in terms of true belief. However, as we saw in the previous chapter, this would, at most, imply that true belief is not non-instrumentally valuable for *everyone*, not that true belief is not non-instrumentally valuable for *anyone*, including those who engage in inquiry, as the latter is construed by the great majority of epistemologists. On the other hand, it could be interpreted as a claim to the effect that there is no such thing as inquiry at all. It is terribly hard to make sense of such a claim, and an argument terminating in such a conclusion would, most plausibly, not be properly described as a *modus ponens* but as a *modus tollens*.

### 3.5. The Trials and Achievements of Understanding

To recapitulate, it has been argued that we lack convincing pre-theoretic reason to take knowing to be of greater epistemic value than merely believing truly, or to consider being justified non-instrumentally valuable. The present section considers the third and final candidate for a bearer of such value: understanding. Two recent accounts of understanding will be discussed, the first one due to Kvanvig, and the second one due to Pritchard. Let us consider the former first.

#### 3.5.1. Objectual Understanding

Kvanvig (2003) is concerned with what he refers to as *objectual* understanding, i.e., the kind of understanding ascribed with respect to an object (as opposed to a proposition), as when we talk about someone understanding politics or the presidency (p. 191). According



to Kvanvig, such understanding is factive (p. 190), and requires the satisfaction of a *coherentist* condition, consisting in “an internal grasping or appreciation of how the various elements in a body of information are related to each other in terms of explanatory, logical, probabilistic, and other kinds of relations that coherentists have thought constitutive of justification” (pp. 192-193). Crucially, however, understanding is, on Kvanvig’s analysis, compatible with luck. As such, an analysis of understanding does not need to satisfy a Gettier condition and is not susceptible to the worries discussed at the beginning of §3.4.

This opens up the possibility that it can be shown for understanding what Kvanvig argues cannot be shown for knowing, i.e., that it is more valuable than any proper subset of its components. However, as stressed above, it is one thing to construct a theory implying that *p*, and quite another to provide pre-theoretic reasons as for why we *should* want our theory to imply that *p* in the first place. According to Kvanvig, “there is at least as much intuitive support for the idea that understanding has value beyond that of its subparts as there is for the idea that knowledge has such value” (p. 188). In light of what has been argued above, this statement might be less comforting than Kvanvig probably intended it—a problem that is exacerbated by the fact that Kvanvig provides no *additional* pre-theoretic support for such value in the case of understanding, beyond that which he assumes already exists for knowing.

Instead, Kvanvig proceeds to argue that this alleged surplus value of understanding can be accounted for. Since the account provided is a highly intriguing one, let us make an exception and waive the above requirement to provide pre-theoretic

support, if only for the sake of the argument. More specifically, let us grant Kvanvig the following dialectical advantage: If the account he provides makes it plausible that

(U) understanding is more epistemically valuable than merely believing truly,

then we have good reason to believe that we have found a candidate counterexample to (2). After all, in analogy with what was said about (K) above, (U) could neither be accounted for with reference to the factivity of understanding, nor in terms of instrumental connections to true belief on part of any components of understanding (on pain of falling prey to the swamping problem). Consequently, the most plausible inference to draw from (U) would be that the surplus value of understanding over mere true belief is due to a fundamental, non-instrumental value, distinct from that which applies to true belief, which, clearly, would contradict (2). On the other hand, if the account provided by Kvanvig does *not* make it plausible that (U) holds—and below it will be argued that it does not—then (a) the same would have to be said for the stronger claim about understanding being more epistemically valuable than *any* proper subset of its components, and (b) (2) remains unchallenged.

According to Kvanvig, the value relevant to accounting for (U) is that associated with the satisfaction of the coherentist condition. In order to avoid the swamping problem, Kvanvig takes the relevant kind of value to be non-instrumental (p. 200). In explicating the relevant kind of value, Kvanvig makes a distinction between intentional and effective means to goals, such as the goal of believing truly (pp. 63-64). Effective means are instrumentally valuable in so far as they raise the likelihood of securing the

relevant ends. Intentional means are also valuable, according to Kvanvig, but not because they raise the likelihood of securing the goals in question. Such means are involved in *trying* to attain goals, and account for the value of such attempts even if they fail to raise the likelihood of attaining the goals in question as a result of trying. It is in this kind of value that we find the surplus value of understanding over mere true belief. More specifically, satisfying the coherentist condition is “valuable because it is constituted by adopting intentional means to the goal of truth” (p. 200).

Why is merely *trying* to attain true belief valuable, even if so doing goes no lengths whatsoever toward actually attaining true belief? Kvanvig’s answer invokes a story not too different from the one alluded to in §3.4, albeit with an introspectionist twist:

Truth is not a property that is always reflectively accessible [...] and so we must adopt some means for identifying beliefs to hold in order to achieve the indirect goal of believing a claim if and only if it is true. So we should try to have, or value, beliefs with some other property, one that we can tell directly and immediately whether a belief has (Kvanvig 2003, pp. 65-66).

Internalist justification is one such property—a “mark of truth” (p. 64), as Kvanvig calls it—albeit one that is unable to account for the surplus value of knowledge over any subset of its constituents due to aforementioned Gettier worries. Another such property is the internal grasping of how various elements in a body of information are related. Moreover, according to Kvanvig, the value associated with adopting such a grasp as an intentional means toward believing truly is exactly what accounts for (U).

Or does it? Consider two possible interpretations of Kvanvig’s claim about the value of adopting intentional means. First, let us assume that the idea is that simply trying

to attain true belief—no matter the instrumental merits of the means utilized—is epistemically valuable. This would constitute a highly implausible commitment. Imagine a subject that sincerely tries to attain true belief, yet opts for a reflectively accessible means toward truth that, due to a particularly deviant conception on the subject’s part about what will actually get her the truth, actually is highly detrimental to her attainment of true belief. Claiming that the mere fact that she has adopted certain intentional goals by virtue of sincerely *trying* to attain true belief makes her conduct epistemically valuable seems highly implausible.<sup>29</sup> This is not to deny that opting for reflectively accessible surrogates for truth might often be a good thing for reasons not too different from those discussed in the previous section, but merely to point out that opting for just *any* reflectively acceptable surrogate does not yield any epistemic value as such.

Second, it might be argued that this interpretation ignores the *nature* of the coherentist condition that Kvanvig takes to be central to understanding. That is, let us assume that Kvanvig has in mind the significantly weaker claim that aiming for truth by way of some *particular* intentional means—including but not necessarily limited to the kind of grasp involved in understanding—is non-instrumentally valuable. In other words, perhaps what explains the surplus value of understanding is not that *something* is intentionally pursued as a mark of truth, but that what is intentionally pursued thus is something with epistemically virtuous properties, where the virtue in question (on pain of rendering the account susceptible to swamping problems) is *not* parasitic upon

---

<sup>29</sup> It might be objected that there is still *something* admirable or valuable about someone sincerely pursuing truth, even if she does a horrible job at it. That may be so—after all, we admire people and their conduct for all kinds of reason, most of which are not importantly related to the formation of true belief. That said, to be relevant presently, the kind of admiration or value at issue would need to be *epistemic*, and that is what seems highly doubtful.

instrumental connections. This interpretation seems more faithful to what Kvanvig eventually concludes about the value of understanding. At the very end of his book, he writes about the grasping of explanatory relationships that is involved in understanding that

[...] to have mastered such explanatory relationships is valuable not only because it involves the finding of new truths but also because finding such relationships organizes and systematizes our thinking on a subject matter in a way beyond the mere addition of more true beliefs or even justified true beliefs. Such organization is pragmatically useful because it allows us to reason from one bit of information to other related information that is useful as a basis for action, where unorganized thinking provides no such basis for inference. Moreover, such organized elements of thought provide intrinsically satisfying closure to the process of inquiry, yielding a sense or feeling of completeness to our grasp of a particular subject matter. In sum, understanding is valuable because it is constituted by subjectively justified true belief across an appropriately individuated body of information that is systematized and organized in the process of achieving understanding, and subjectively justified true belief that is systematized in this way is valuable (Kvanvig 2003, p. 202).

In other words, there are three potential sources for the value of understanding. The first one is that the relevant grasp may lead to new truths being found. However, considering that the value in question would be *instrumental*, it could not account for (U), due to aforementioned swamping problem. The second source is that understanding brings systematicity to our thinking. That may be a good thing, but Kvanvig explicitly states that the kind of value relevant to such systematicity is *pragmatic*, not epistemic. As such, it cannot account for the surplus value relevant to (U). The third source is that understanding brings a feeling of closure. Such a feeling may be pleasurable and, hence, *prudentially* valuable. However, as such, it does not seem relevant to epistemic value, nor, consequently, to (U). Hence, even waiving the requirement to provide pre-theoretic support for (U), we still lack reason to believe that the value in question is not exhausted

by (a) the non-instrumental value derived from its factivity, and (b) the instrumental value it may have as an effective means to truth. As such, Kvanvig's notion of understanding poses no challenge to EVM.

### 3.5.2. Understanding-Why

There are, of course, other types of understanding besides objectual understanding. More recently, Pritchard (forthcoming) has defended a view about *understanding-why*, i.e., the kind of understanding that is involved in understanding why such-and-such is the case. Pritchard's account is meant to account for what he refers to as the *distinctive* value of such understanding. To be of distinctive value, Pritchard explains, is both a matter of degree and kind (p. 13). More specifically, while Pritchard takes understanding-why (henceforth, simply "understanding") to be factive (p. 113), he does not take the value of understanding to be exhausted by the value that it has in virtue of its factivity. For understanding to be of distinctive value is for understanding to be both *more* valuable than anything that falls short of understanding, including true belief, and for that surplus value to be due to understanding being of *fundamental* epistemic value.

Presently, we will focus on the latter claim, since it is the more basic one of the two. After all, the surplus value in question is supposed to be explained with reference to the non-instrumental value of understanding. But before delving into Pritchard's theory, let us consider what pre-theoretic reasons we have for at all taking understanding to be distinctively valuable. According to Pritchard, "[t]he intuition that understanding is distinctively valuable is surely even stronger than the intuition that knowledge is distinctively valuable" (p. 101). If what has been argued above is correct, it better be. But

since Pritchard does not elaborate on the claim, let us simply grant him an advantage analogous to that granted to Kvanvig above, and waive the requirement on pre-theoretic support; all that is required for Pritchard's theory of understanding to amount to a challenge to (2) is that it makes plausible that understanding is of a non-instrumental value that does not derive from its factivity.

Pritchard's case for understanding being valuable thus rests on a distinction between two kinds of value that we familiarized ourselves with already in the above, i.e., final value and fundamental value. According to Pritchard, a fundamental epistemic good is an "epistemic good whose epistemic value is at least sometimes not simply [an] instrumental value relative to a further *epistemic* good" (p. 19). More specifically, Pritchard takes the defining feature of fundamental epistemic value to be that it can "act as the terminus for the instrumental regress of epistemic value" (p. 19). Consequently, Pritchard's fundamental goods correspond to what we have been referring to as goods of fundamental (non-instrumental) epistemic value above. After all, any good whose non-instrumental epistemic value derives from the value of one of its components can not act as the relevant kind of terminus, since any question about why *that* good is of non-instrumental epistemic value would have to be answered with reference to a *further* epistemic good, i.e., the epistemic good of the component in question.

Next, consider final value. According to Pritchard, a good is of final value if it is of non-instrumental value *simpliciter*, i.e., if its value holds independently of any considerations about conduciveness whatsoever. As noted in §3.1, assuming that at least some instances of true belief are of fundamental epistemic value is compatible with maintaining that (those) instances are only valuable thus in so far as they pertain to

questions posed by some relevant set of inquirers. In other words, maintaining that some instances of true belief are of fundamental non-instrumental *epistemic* value is compatible with maintaining that those instances are valuable thus only in so far as they are of instrumental non-epistemic value, and, hence, not of non-instrumental value *simpliciter*. So, again, fundamental non-instrumental epistemic value does not imply final value.

However, according to Prichard, “final value *simpliciter* does entail fundamental epistemic value.” “After all,” he writes, “if an epistemic good has a value which is not an instrumental value then, *a fortiori*, it has a value which is not an instrumental epistemic value either” (p. 20). That seems right. But surely it does not go to show that final value implies non-instrumental epistemic value. There are more ways of *not* being of instrumental epistemic value than to be of non-instrumental epistemic value, including to not be of any epistemic value at all. As we noted already in §3.1, there are plausible candidates for finally valuable goods, such as the value that pertains to undergoing an episode of pleasure, that, clearly, are *not* thereby of epistemic value, let alone of non-instrumental epistemic value.

This presents a problem for Prichard, since he rests his entire case for the value of understanding on this alleged implication. More specifically, he argues that the reason that such understanding is of non-instrumental, epistemic value is that it involves a cognitive *achievement*, where an achievement is a success that is due to the achiever’s skills, and thus creditable to the exercise of her abilities. Moreover, according to Prichard, such achievements are finally valuable. How so? Prichard asks us to imagine the following scenario:



Imagine [...] that you are about to undertake a course of action designed to attain a certain outcome and that you are given the choice between merely being successful in what you set out to do, and being successful in such a way that you exhibit an achievement. Suppose further that it is stipulated in advance that there are no practical costs or benefits to choosing either way. Even so, wouldn't you prefer to exhibit an achievement? And wouldn't you be right to do so? If that's correct, then this is strong evidence for the final value of achievements (Pritchard forthcoming, p. 44).

Combining Pritchard's points about understanding and cognitive achievements, his case for the non-instrumental epistemic value of understanding can be reconstructed thus:

- (1) Cognitive achievements are finally valuable.
- (2) All finally valuable goods are of non-instrumental epistemic value.
- (3) Cognitive achievements are of non-instrumental epistemic value.
- (4) Understanding is a cognitive achievement.
- (5) Understanding is of non-instrumental epistemic value.

This is a valid argument. But it is not sound, and for two reasons. First, as called attention to above, all finally valuable goods are *not* of non-instrumental epistemic value, since some finally valuable goods are of no epistemic value at all. This undermines (2).

Second, even setting this concern aside, (1) is independently questionable.

Remember, something is finally valuable if and only if it is non-instrumentally valuable without qualification. If we grant Pritchard's claim that we would prefer to exhibit a cognitive achievement in the scenario imagined (as opposed to, perhaps, having no preference either way), we have some reason to rule out exhibiting an achievement being of instrumental *practical* value. However, this establishes neither that (a) exhibiting an

achievement is of *non-instrumental* practical value (since there are more ways of not being of instrumental practical value than to be of non-instrumental practical value, including to not be of any practical value at all), nor that (b) exhibiting an achievement is not of some non-instrumental *qualified* value. As for the latter, it might, for example, be that there is something intrinsically pleasing about exhibiting an achievement, and that such achievements, hence, may be of non-instrumental *prudential* value. However, none of this serves to establish the far stronger claim that cognitive achievements are of final value, i.e., of non-instrumental value *without qualification*.

By way of conclusion, Pritchard's account of understanding does not successfully support the claim that such understanding is of non-instrumental epistemic value, even if we waive the requirement to provide pre-theoretic support. For all we are being told, the value of understanding is exhausted by the non-instrumental value that derives from its factivity, and whatever value is associated with understanding being a cognitive achievement. However, contrary to what Pritchard argues, it is not clear that the latter kind of value is final. Moreover, even if it is granted that cognitive achievements are finally valuable, and that understanding is a form of cognitive achievement, this does not imply that understanding is of non-instrumental *epistemic* value. As such, we may conclude that Pritchard's account of understanding poses no problem for (2), nor consequently for EVM.

### 3.6. Conclusion

Recently, epistemic value monism has come under attack by philosophers arguing that we cannot account fully for the domain of epistemic value in monistic terms. But as we have seen, the relevant arguments fail to establish any such thing. For one thing, there is a presumption of monism due to considerations about axiological parsimony. Granted, such a presumption would be defeated by positive evidence to the effect that the relevant kind of monism makes us unable to fully account for the domain of epistemic value. However, a proper examination and subsequent rebuttal of the most promising counterexamples to monism, in the form of three candidates for states of fundamental, non-instrumental epistemic value, casts serious doubt upon the claim that there is any such evidence. Consequently, epistemic value monism still stands.

## CHAPTER 4

### A FRAMEWORK FOR EVALUATION

The previous two chapters established two things. First, believing truly is of non-instrumental epistemic value (albeit not necessarily for everyone), in virtue of true belief being a goal of inquiry. Second, a careful consideration of the most promising counterexamples to epistemic value monism suggests that believing truly is unique in being valuable thus. As noted earlier, this is not to say that *all* instances of true belief are non-instrumentally valuable. As will be argued below, only the subset consisting of true beliefs that are *significant* are valuable thus. Much of §4.1 will be devoted to spelling out the relevant notion of significance, by explaining how believing some true propositions is in some relevant epistemic sense more *important* to inquirers than believing other true propositions, specifically in so far as the truths in question serve to answer questions that the inquirers at issue are posing.

As such, the account to be developed tells us something about what inquirers *should* do, in so far as they want answers to their questions: they should employ the particular practices that will enable them to answer those questions. §4.2 will be concerned with tracing the limits of this ‘should.’ In particular, it will be suggested that it follows from the account to be defended that there are very real limits to the extent to which we can make *epistemic* judgments about what questions inquirers should pose (as opposed to about what methods or practices to opt for, *given* a set of questions).

However, it is also argued that, rather than indicating a flaw in the theory, this simply

suggests that it is able to maintain a clear distinction between epistemic and other kinds of betterness (e.g., moral, practical, etc.), and, thereby, account for the specialized nature of epistemic evaluation.

#### 4.1. Defining Epistemic Betterness

As was noted already in chapter 1, one of epistemology's main missions is to evaluate and improve our epistemic ways through ameliorative recommendations. Such recommendations can be implemented either on the individual or on the social level. For reasons that will be discussed at length in the next chapter—in essence, that ameliorative advice provided directly to the individual rather than implemented by way of a social paternalistic mandate are likely to face problems of defection—the focus will presently be on the evaluation and amelioration of social practices. By a 'social practice,' I mean a coordinated implementation or utilization of a method or a set of methods by a group of inquirers, where a 'method' is to be understood as an acquired (as opposed to native) belief-forming process.<sup>1</sup>

However, before we can talk of an *improvement* of such practices, we need to talk about what it would mean for one practice to perform *better* from an epistemic point of view than another. More specifically, any evaluation (and attempt at amelioration) needs to be made against the background of (a) an account of what practices need to produce in order to be positively evaluated, as well as (b) a framework for comparing practices with respect to the production of the relevant goods. In short, we need a framework of

---

<sup>1</sup> See Goldman (1986, p. 93).

evaluation that involves an account of *epistemic betterness*, as applied to practices. The present chapter develops such a framework in terms of the monistic account of epistemic value defended in the previous two chapters.

How are we to spell out the notion of epistemic betterness, given such an axiology? To a first approximation, we may try to do so with reference to the *reliability* of a practice, i.e., the ratio of true to false belief that the practice produces (or *would* produce) under a given set of circumstances. However, a practice may generate a high ratio but a very low *number* of true beliefs. For this reason, we may try to invoke a combination of reliability and *power*, where the latter corresponds to the number of true belief produced, given a set of circumstances. While an improvement, a further problem arises, to the effect that a practice may be extremely reliable and powerful without answering any questions whatsoever, as in a case where it only generates high ratios and numbers of true belief that are of no relevance whatsoever to the concerns of the agent.<sup>2</sup>

A more promising strategy is, therefore, to specify the relevant notion of power further in terms of *question-answering power*.<sup>3</sup> In other words, what is epistemically relevant as far as power is concerned, is not that the agent arrives at a large number of true beliefs *per se*, but that she arrives at a large number of true beliefs that answer questions she wants answered.<sup>4</sup> In light of this, consider the following triad of definitions, providing a first stab at what constitutes epistemic betterness, as it applies to practices:

---

<sup>2</sup> Cf. Alston (2005) and Goldman (1986).

<sup>3</sup> See Goldman (1986, p. 123).

<sup>4</sup> This is a slight departure from Goldman's formulation in terms of "the *proportion* of questions [a cognitive system] wants to answer that it can answer (correctly)" (1986, p. 123; emphasis added). (See Goldman 1992a, p. 195 for a similar formulation in terms of practices.) However, note that a formulation in terms of proportions of questions that the subject wants answered would imply that a practice can be rendered less powerful (for

- (B) A practice  $P_1$  is *epistemically better* than another practice  $P_2$  for a subject  $S$  iff  $P_1$  outperforms  $P_2$  for  $S$ , where
- (O) a practice  $P_1$  *outperforms* another practice  $P_2$  for  $S$  iff  $P_1$  yields (i) a higher ratio and number, (ii) an equal ratio but a higher number, or (iii) a higher ratio but an equal number of beliefs that are significant to  $S$ , and
- (S<sub>1</sub>) a belief is *significant* to  $S$  iff it answers questions on  $S$ 's research agenda.

For ease of reference, I will henceforth refer to any combination of definitions by way of their abbreviations, together with a '+.' For example, I will refer to the combination of above definitions as '(B)+(O)+(S<sub>1</sub>).' (Definitions clearly assumed but not at issue will be suppressed for the sake of brevity.) The above triad of definitions raises a number of questions that will be addressed in the remainder of the section, where they will also be refined and qualified in a number of ways. First off: What is a research agenda?

#### 4.1.1. Questions and Informative Answers

A research agenda is, to a first approximation, a non-ordered set containing those questions the pursuit of which the relevant inquirers consider most worthy of their limited epistemic resources. A couple of comments are at place. First, the reasons for specifying the set in terms of the questions deemed *most* worthy of pursuit will be discussed in due course (see §4.2.1), and may be put to the side presently. Second, the questions featured

---

that subject) simply by increasing the number of questions *asked*, while a formulation in terms of numbers would be sensitive to the factor that seems more relevant, i.e., the numbers of questions *answered*.

on the agenda do not need to be *explicitly* formulated by the inquirer. A research agenda may simply be the most plausible reconstruction of the questions that the inquirer deems most pressing, as revealed by the choices that she makes in allocating her resources.

Third, while there is, clearly, legitimate room for a notion of an agenda as applicable to an inquirer because an institution or other pertinent body considers the relevant set of questions worthy of her limited resources, I will reserve the discussion of such an extension of the present notion for another occasion.<sup>5</sup>

For present purposes, we may assume that there are two broad classes of reasons for pursuing some questions rather than others. On the one hand, there is the class of reasons that we may ascribe to *curiosity*, or a desire for an answer to a question for the mere purpose of attaining that answer. I will not delve into the nature or source of this desire, but merely take it as a psychological fact that there is such a thing as a “sheer intellectual curiosity” (Hempel 1965, p. 333) that, to paraphrase Larry Laudan (1977, p. 225), seems to be every bit as compelling to many of us as our need for clothing and food. On the other hand, there is also the class of reasons that we may ascribe to a desire to attain answers for some *other* sake than the mere attainment of those answers, be it for moral, practical, or some other kind of reason.

However, what is important to note is that, regardless of whether an inquirer pursues answers to questions because of curiosity or for some other reason, she will *not* necessarily consider every true answer equally worthy of pursuit. To illustrate this point, consider two social practices,  $P_1$  and  $P_2$ .  $P_1$  generates  $x$  answers to questions on the relevant inquirer’s agenda at a truth ratio of  $y$  within the relevant circumstances, while the

---

<sup>5</sup> See Goldman (1999, p. 94) for a relevant discussion.



corresponding values for  $P_2$  are  $x + n$  and  $y$ , where  $x$ ,  $y$ , and  $n > 0$ . Moreover, while  $P_1$  generates at least some informative answers,  $P_2$  exclusively generates answers that constitute utterly *uninformative* answers by all plausible standards. Still, since (B)+(O)+(S<sub>1</sub>) makes no distinction between informative and uninformative answers, we would have to say that  $P_2$  is epistemically better than  $P_1$  (for the inquirer in question). Needless to say, that would not make for particularly good advice from the point of view of amelioration.

I do not have a theory about what constitutes sufficiently informative answers, but I am also not so sure that we need one. To make this point, I will, first, consider an illustration of the relevant notion and, then, suggest that we have a sufficiently good intuitive sense of paradigmatic cases of uninformative and sufficiently informative answers, respectively, for us to not require a theory. By way of illustration, consider the following question:

(Q) “What caused  $x$ ?”

We can imagine one answer that lists the actual cause(s) of  $x$ . Then we can imagine another answer: “Something.” Both answers are true and on topic, but there is a sense in which the latter is genuinely unhelpful. Returning to the two classes of reasons for pursuing answers to questions introduced above, this might be so for either of two reasons. Assuming that we are interested in finding an answer to (Q) for some other reason than mere curiosity, e.g., if  $x$  constitutes a malfunction that we want to correct and prevent in the future, an answer in terms of “something” is unhelpful because it tells us

nothing whatsoever about how to go about doing this. Similarly, if we are interested in (Q) as a result of curiosity, the mere fact that we are simply curious about an answer to (Q) does not imply that we would accept just *any* true answer as a sufficiently informative answer. For example, someone pondering what gave rise to some ancient event may do so without having any practical (or moral, or prudential, or what have you) stake in the answer—maybe she just finds thinking about these kinds of things intellectually stimulating. Still, her curiosity might, nevertheless, only be satiated given that the answer is sufficiently informative. Indeed, had she *not* been curious about the question, she might have settled for a completely uninformative answer, like “something.”

In other words, the intuitive notion of an answer being sufficiently informative is roughly that of the answer actually *answering* the question, in a sense that goes beyond saying something that is both true and on topic. Interesting formal work has been done on specifying the exact features that answers are typically required to have (beyond that of being true and on topic) for a variety of question types.<sup>6</sup> However, more relevant than the details of the relevant literature—details that, for present purposes, would add formal detail but not necessarily explanatory value—is the kind of data invoked in the literature, i.e., intuitive judgments of ordinary speakers regarding what answers constitute sufficiently informative answers, given a particular question. In line with this methodology, I will characterize what it is for an answer to be sufficiently informative in operational terms as follows:

---

<sup>6</sup> See, e.g., Belnap and Steel (1976), as well as Jaworski (2009) for a recent application of Belnap and Steel’s framework to certain kinds of how-questions.

- (A) an answer  $a$  is a *sufficiently informative* answer to a question  $q$  iff (i)  $a$  is true, and (ii) any normally functioning person that grasps the contents of both  $a$  and  $q$  would accept  $a$  as an answer to  $q$ , provided that she takes  $a$  to be true.

This might be more of a clarification than a definition. For one thing, a completely satisfactory definition would need to spell out the relevant notions of normal functioning, content and what it is to grasp content. That being said, I believe that (A), nevertheless, is sufficient for present purposes, given our intuitive ability to provide reasonable verdicts at least about paradigmatic instances of sufficiently and insufficiently informative answers, respectively. Moreover, against the background of (A), we may reformulate (S<sub>1</sub>) as follows:

- (S<sub>2</sub>) A belief is *significant* to  $S$  iff it constitutes a sufficiently informative answer to questions on  $S$ 's research agenda.

(B)+(O)+(S<sub>2</sub>)+(A) yields the verdict that  $P_1$  is epistemically better than  $P_2$  in the above example, and, thereby, constitutes an improvement on (B)+(O)+(S<sub>1</sub>).

#### 4.1.2. Problems of Individuation

(O) invokes the notions of ratios and numbers of belief. This seems to be presupposing that beliefs can be individuated in systematic ways. However, any given belief can be described at a number of degrees of generality. For example, should my belief “The sun

is shining right now” be distinguished from my belief that “The sun is shining on September 14<sup>th</sup>, 2009”? Or from my belief that “The sun is shining in New Brunswick on September 14<sup>th</sup>, 2009”? It is hard to say. More than that, in want of a principled account of what degree (or degrees) of generality should be privileged, there is a potential problem about *double* counting and about *not* counting what should be counted. For this reason, one may have one’s doubt about the usefulness of talking about ratios and numbers of belief. These doubts may be intensified by the fact that we, on the present account, need to individuate ratios and numbers of *significant* (true) beliefs, i.e., beliefs that constitute acceptable answers to questions on the relevant agenda(s). Consequently, we not only need to individuate beliefs, but also answers and agenda items.

In response to these worries, it should be noted that talk about ratios and numbers, albeit convenient for the purposes of thinking clearly about the relevant issues, has as Andrew Latus (2000) puts it a “misleading air of precision,” as far as the evaluation of actual practices is concerned. More specifically, what I would like to suggest is that epistemic evaluation—at least as it concerns us here—does not require a fully systematic account of belief, answer, and agenda item individuation. It only requires that we can approximate the relevant magnitudes to an extent sufficient for making informative and roughly accurate comparative judgments.

Let us start with the individuation of belief. As just noted, the kinds of claims we have to be able to make on the present account are comparative claims. More than that, the kind of comparative claims that concern us here are explicitly *ordinal* claims. Consequently, when comparing two practices, the only thing required is that we are able to make comparative claims about one practice yielding *more* significant true beliefs than

another, or one practice yielding a greater *preponderance* of significant true belief than another. What I would like to suggest is that the kinds of problems coming out of dangers of double-counting and not counting what should be counted are *not* on a scale that undercuts the possibility of making such ordinal claims. And given that we do not need anything beyond such claims to make sense of the comparisons involved in (O), the present account does not fall prey to a problem of individuation, as far as the individuation of beliefs is concerned.

A similar point can be made in relation to the individuation of answers. It was suggested above that we have a fairly robust intuitive sense of what constitutes sufficiently informative answers. Moreover, as already stressed, the complications associated with individuation would have to be severe enough to make impossible even informed *approximations* in the form of comparative, ordinal claims about the relative merits of two practices. In light of this, an answer similar to the one given above in relation to belief individuation may be given also in the case of answer individuation. More specifically, I submit that we can approximate the relevant magnitudes and preponderance of sufficiently informative answers to an extent sufficient for making informative and roughly accurate comparative judgments, even in the absence of a fully systematic account of how to individuate acceptable answers.

Consider, finally, the individuation of agenda items. Any given inquirer may want a whole host of different questions answered—questions that, moreover, may be more or less interconnected and of different degrees of generality. For example, consider a cancer researcher. At the most general level, a cancer researcher may be interested in the causes, treatment and/or prevention of cancer. With respect to each general area, the researcher

may, in turn, be interested in a series of more specific questions. For example, with respect to the causes of cancer, the researcher may be interested in identifying common types of mutated genes that tend to be causally implicated in turning normal cells into cancer cells. Beyond that, the researcher may be particularly interested in a specific subset of such genes, perhaps because understanding the nature of that sub-set has proven to be particularly important in understanding common forms of tumor progression, which, in turn, may have implications for the success of different forms of treatment and strategies for prevention. So, how many items figure on her agenda?

It goes without saying that any numerical judgment made in this context is going to be largely arbitrary. Still, it is likely that different disciplines have developed certain rough standards for how to group different questions into hierarchies—hierarchies that show up in everything from grant applications to conference themes, anthologies, and calls for papers. This is not to suggest that these hierarchies are, in any way, cutting the space of taxonomical possibilities at its natural joints (if there even are any such natural joints). Still, it is likely that we can expect a certain degree of *homogeneity* within disciplines with respect to how inquirers will characterize their agendas. Moreover, if we can expect there to be such hierarchies, and we see to it that we focus on the *same* level of any given hierarchy in comparing how well two practices do with respect to answering the questions contained on that level, then the possibility that every form of inquiry can be described with reference to a great multitude of different questions does not seem to threaten this particular use of the notion of a research agenda in evaluation.

### 4.1.3. Shared Agendas

So far, we have been concerned with what it is for one practice to be epistemically better than another *for an individual*. As for what it is for one practice to be better than another for a *group* of individuals, consider the following reformulations of (B), (O) and (S<sub>2</sub>):

- (B<sub>G</sub>) A practice  $P_1$  is *epistemically better* than another practice  $P_2$  for a group of subjects  $G$  iff  $P_1$  outperforms  $P_2$  for  $G$ , where
- (O<sub>G</sub>) a practice  $P_1$  *outperforms* another practice  $P_2$  for  $G$  iff  $P_1$  yields (i) a higher ratio and number, (ii) an equal ratio but a higher number, or (iii) a higher ratio but an equal number of beliefs that are significant to the subjects in  $G$ , and
- (S<sub>G</sub>) a belief is *significant* to  $G$  iff it constitutes a sufficiently informative answer to questions on  $G$ 's shared research agenda.

As these definitions make clear, what it is for one practice to be epistemically better than another for a group is just a matter of aggregating the relevant ratios and numbers for the group's individuals, on the crucial assumption that the individuals in question share a research agenda. In the interest of simplicity of exposition, I will stick to talking in terms of epistemic betterness for an individual subject, and leave to the reader to draw the analogous implications for groups of subjects in accordance with (B<sub>G</sub>)+(O<sub>G</sub>)+(S<sub>G</sub>), in contexts where she deems that appropriate.

However, note that none of this is to suggest that we cannot imagine comparative epistemic measures across agendas. For example, a measure that can be extracted from

the present account fairly easily is the following: A practice  $P_1$  is more *robust* than another practice  $P_2$  iff  $P_1$  outperforms  $P_2$  for a wider range of agendas than the range of agendas within which  $P_2$  outperforms  $P_1$ . Note, however, that one practice being more robust than another does not imply that the former is epistemically better than the latter, for the simple reason that epistemic betterness is defined with reference to a *single* agenda. This mirrors the intuition that a practice might be epistemically better than another given a very specific set of questions, without, thereby, being successfully applicable to a whole host of other questions, as is surely the case with several practices within the sciences that involve very specialized forms of methods, practices, and instruments. We will return to this issue briefly in §5.3 below. Beyond that, I will not have anything to say about comparative epistemic measures that may apply across agendas.

#### 4.1.4. On Incomparability

On (O), one practice outperforms another if and only if the former yields (i) a higher ratio and number, (ii) an equal ratio but a higher number, or (iii) a higher ratio but an equal number of true belief that are significant to  $S$  than the latter practice does. But what about scenarios in which none of (i) through (iii) holds? By mapping out all the nine (i.e.,  $3^2$ ) possibilities, we see that there are easy as well as hard cases. More specifically, consider the following table, representing the comparison of one practice,  $P_1$ , with another practice,  $P_2$ , where ‘+’ represents  $P_1$  receiving a higher score than  $P_2$  on the relevant factor, ‘-’ represents  $P_1$  receiving a lower score than  $P_2$  on the relevant factor, and ‘=’ represents  $P_1$  receiving the same score as  $P_2$  on the relevant factor:



	REL	PWR	
1.	+	+	B
2.	=	+	B
3.	+	=	B
4.	=	=	E
5.	+	-	I
6.	-	+	I
7.	=	-	W
8.	-	=	W
9.	-	-	W

Rows 1 through 3 cover cases in which (i), (ii), or (iii) hold, and where  $P_1$ , consequently, is epistemically better than  $P_2$ . Row 4 covers a case in which  $P_1$  and  $P_2$  are *equally* epistemically good, and rows 7 through 9 cases in which  $P_1$  is epistemically *worse* than  $P_2$ , given the following two definitions:

- (E) Two practices,  $P_1$  and  $P_2$ , are *equally* epistemically good for a subject  $S$  iff  $P_1$  and  $P_2$  yield equal ratios and numbers of beliefs that are significant to  $S$ .
- (W) A practice  $P_1$  is epistemically *worse* than another practice  $P_2$  for a subject  $S$  iff  $P_1$  yields (i) an equal ratio but a lower number, (ii) a lower ratio but an equal number, or (iii) a lower ratio and a lower number of beliefs that are significant to  $S$  than  $P_2$  does.

This leaves us with rows 5 and 6, involving cases where one of the two practices fares better on one score, while the other fares better on the other score. What are we to say about these cases? Consider the following:

- (I) Two practices,  $P_1$  and  $P_2$ , are epistemically *incomparable* for a subject  $S$  iff neither (B), (E) or (W) holds between  $P_1$  and  $P_2$  for  $S$ .

This way of understanding incomparability is fairly standard in the literature.<sup>7</sup> On (I), the cases covered by rows 5 and 6 are instances of epistemic incomparability. That is, while one practice is epistemically better with respect to power, and one with respect to reliability, there is no additional covering consideration in light of which it can be inferred that one is epistemically better than the other *simpliciter*. Consequently, in these cases of incomparability, there are no *epistemic* reasons to prefer one of the practices to the other.

However, it should be noted that this is perfectly compatible with there being a whole host of *non-epistemic* reasons to prefer one practice over another, even in a scenario where they are incomparable from an epistemic point of view. For example, we can easily imagine cases in which avoiding error is more important than arriving at many true answers, due to the fact that the moral or practical costs of a *mistaken* judgment about the matters at hand far outweighs the benefits of being able to make *a* judgment. In such cases, reliability might trump power, and  $P_1$ , hence, be preferable on non-epistemic grounds to  $P_2$  in the scenario represented by row 5, despite the two practices being epistemically incomparable. Similarly, we can imagine cases in which power trumps reliability, as in contexts of intellectual exploration or speculation where it is more

---

<sup>7</sup> See, e.g., Chang (1997). This is not to say that the idea that this is the proper way to understand incomparability is unchallenged. For example, Chang (2002) argues for the existence of a fourth relation (beyond better, worse, and equally good), i.e., *parity*—a relation that I will not consider presently.

important that an answer is generated than that the proportion of false answers to total answers generated is minimized. In such cases,  $P_1$  would be preferable on non-epistemic grounds to  $P_2$  in the scenario represented by row 6.

Moreover, such paradigmatically non-epistemic considerations can not only provide relevant covering considerations in cases of epistemic incomparability, but also in cases of epistemic comparability. In the latter cases, however, such considerations are not invoked in the *absence* of any epistemic reasons to opt for one practice over the other, but in the presence of epistemic reasons that are *overridden* by non-epistemic considerations. Still, *qua* non-epistemic considerations, they fall outside the scope of the present investigation, beyond the extent to which they may serve to account for the intuition that (a) epistemic incomparability is compatible with having reasons (as in: *non-epistemic* reasons) to prefer one practice over another, and that (b) the epistemic reasons present in cases of epistemic comparability may be overridden by non-epistemic considerations. We will find reason to return to the latter point in the next chapter (in §5.4).

#### 4.2. On the Specialized Role of Epistemic Evaluation

Let us take stock at what has been argued so far. The account of epistemic betterness developed above may be summed up as follows:

- (B) A practice  $P_1$  is *epistemically better* than another practice  $P_2$  for a subject  $S$  iff  $P_1$  outperforms  $P_2$  for  $S$ , where

- (O) a practice  $P_1$  *outperforms* another practice  $P_2$  for a subject  $S$  iff  $P_1$  yields (i) a higher ratio and number, (ii) an equal ratio but a higher number, or (iii) a higher ratio but an equal number of belief that are significant to  $S$ , and
- (S<sub>2</sub>) a belief is *significant* to  $S$  iff it constitutes a sufficiently informative answer to questions on  $S$ 's research agenda, where
- (A) an answer  $a$  is a *sufficiently informative* answer to a question  $q$  iff (i)  $a$  is true, and (ii) any normally functioning person that grasps the contents of both  $a$  and  $q$  would accept  $a$  as an answer to  $q$ , provided that she takes  $a$  to be true.

On this account, epistemic betterness is defined for practices, *given* a research agenda. The account, thereby, tells us something about what inquirers *should* do, in so far as they want to arrive at belief that answer their questions. In the remainder of this chapter, I will address two concerns. First, while answering any questions on our agendas, clearly, might be a good thing from an epistemic point of view, might it not be epistemically *better* to answer some questions rather than others? In particular, might it not be epistemically better to answer questions that we are *more interested* in than questions that we are less interested in (*ceteris paribus*)? Second, what epistemic judgments can we make with respect to what items should figure on agents' agendas in the first place? Let us consider each question in turn.

#### 4.2.1. Degrees of Interest

According to (S<sub>2</sub>)+(A), it is a necessary and sufficient condition on a true belief being significant to an inquirer that the belief (or, to be exact: its content) constitutes a sufficiently informative answer to questions on her agenda. As we have seen, this makes significance *agent-relative* in the sense that what true beliefs are significant is, in part, a function of the questions posed by the inquirer. At the same time, significance is on the present account an *objective* matter, at least in the following sense: A belief  $b_1$  is *not* necessarily more significant for an agent than another belief  $b_2$  simply because she considers the questions answered by  $b_1$  *more interesting* than the answered by  $b_2$ .

This might be taken to present a problem. To see why, consider a scenario discussed recently by Don Fallis (2006), in which there are two practices,  $P_1$  and  $P_2$ , available to an agent.<sup>8</sup>  $P_1$  yields as its only output one believed proposition that answers one question that the inquirer takes great interest in, *viz.* the question of how the world could be made a good and happy place, while  $P_2$  yields as its only output two believed propositions that answer two questions the inquirer takes some (but less) interest in, *viz.* the questions when *Hill Street Blues* and *Streets of San Francisco* come on TV, respectively. Assuming that the inquirers take *sufficient* interest in all three questions for them to be featured on her research agenda, and that other things are equal—e.g., that the believed propositions answer no *more* than one question each—it follows on the present account that  $P_2$  is epistemically better for the subject than  $P_1$ .

---

<sup>8</sup> Fallis formulates his argument in terms of sets of doxastic states rather than in terms of the processes or practices that generate these sets. In order to make his argument applicable to the present account, however, I have reformulated it in terms of practices.

Is this a problem? To get a better grip on the issues relevant to answering this question, let us consider Fallis' suggestion as for an alternative account, on which the epistemic value of forming a true belief is weighted by the degree of interest that the agent has in the truth in question.<sup>9</sup> Worries about how to properly quantify degrees of interest aside, Fallis' framework would, thereby, enable us to say that  $P_1$  is epistemically better than  $P_2$ , due exactly to the fact that the inquirer deems the question answered by way of  $P_1$  more interesting than the questions answered by  $P_2$ . Fallis introduces his account—as well as the above scenario—in a criticism of Goldman (1999). According to Goldman, epistemic evaluation is restricted to epistemic performances that pertain to questions that interest the inquiring agent, in much the same way as epistemic evaluation on the present account is restricted to questions that are featured on the relevant inquirer's research agenda.<sup>10</sup> As such, the *presence* of interest, clearly, plays a role on Goldman's account, even though *degree* of interest does not.

Goldman opts for an account in terms of the presence as opposed to degree of interest in an attempt to balance two desiderata. On the one hand, he acknowledges that “a good epistemic practice should not be indifferent to the relative amounts of interest that a given agent takes in different questions” (Goldman 1999, p. 95). On the other hand, he wants to settle interest with a *modest* as opposed to a *pervasive* role in evaluation, where the latter “might go so far as to use the degree of interest in a question to give an overall weighting to an agent's veritistic [i.e., truth-related] success” (p. 95). Goldman sees two problems with letting interests play a pervasive role in evaluation:

---

<sup>9</sup> See Fallis (2006, pp. 181-182).

<sup>10</sup> See Goldman (1999, pp. 88 and 94-96).

First, one person's questions of interest might be easier to answer than those of a second. The fact that the first scores higher on the set of questions that interest *her* does not demonstrate superior intellectual skill. Second, the intensity of interest a person takes in a question may reflect factors that do not properly belong in an epistemological analysis. The magnitude of interest may depend on the person's financial or mortal stake in the question. But that does not mean that his succeeding or failing to answer those questions correctly should loom disproportionately large in an epistemic analysis. This would lead down the path toward abandoning the specialized, veritistic mission of epistemology in favor of a more purely pragmatic enterprise (Goldman 1999, p. 95).

However, note that Goldman's two objections to providing degrees of interests with a pervasive role in epistemic evaluation apply even if reformulated in terms of the *presence* (as opposed to degrees) of interest. As for easy answers, it is sufficient that a person only is interested in easy questions for an epistemic evaluation in terms of the generation of true belief to be skewed in her favor, when compared to someone that is only interested in hard questions. (This issue will be discussed in more detail below.) As for intensity of interest, just like a person's *magnitude* of interest may depend on the her financial or mortal stake in the question, so, too, for her having *any* interest in the question. Consequently, it does not seem that Goldman's objections serve to favor giving degrees of interest a modest rather than a pervasive role—which is, I take it, exactly Fallis' point.

Where does this leave the present account? Unlike Goldman, I have no quarrel with Fallis' account *per se*. However, within the context of the present framework, weighting the significance of answering agenda items by the degree of interest in the corresponding question would not so much amount to an unwelcome as to a somewhat unnecessary feature. The reason is that we do not imagine the typical research agenda as one that contains *both* (what tends to be considered) paradigmatically pressing concerns, such as those pertaining to the variety of projects that concern the scientific and political

community, and questions about when different TV shows come on TV. More to the point, we are imagining that the span of degrees of interest that we have for the different items on a particular agenda is far more narrow than in the kind of scenario Fallis imagines, for the simple reason that what ends up on the research agenda are the questions we deem *most* interesting out of all the questions that we may choose to pursue. Consequently, it is not that degree of interest plays no role in the present account. It does, but rather than playing a role by way of weighting, it plays a role in that agenda items correspond to those—and only those—questions the pursuits of which we consider most worthy of our limited epistemic resources.

#### 4.2.2. Banal and Harmful Pursuits

Understanding agendas in terms of that which we consider most worthy of pursuit, however, should not be taken to put any substantial restrictions on *what* inquirers may deem most worthy of pursuits. So, let us turn to the question of what epistemic judgments we can make with respect to what items *should* figure on agents' agendas. For one thing, we can often provide people with epistemic reasons to pursue at least some *challenging* questions, if doing so would serve to sharpen their cognitive tools and, thereby, enable them to answer whatever questions already figure on their agenda with a greater efficiency. For another, we can often also make a case for engaging with *specific* questions (be they challenging or not), since doing so may enable us to address already existing agenda items in more productive ways. By the same token, failing to ask certain new questions may have (comparatively) detrimental effects on the extent to which the inquirer will be able to answer questions already on her current agenda. In such cases,



the inquirer has straightforwardly epistemic reasons for expanding or otherwise revising parts her agenda, in so far as doing so will make her epistemically better off with respect to answering the questions she deems most worthy of pursuit.

But what normative, epistemic judgments (if any) can we make beyond this? We may use this question to put pressure on the point made above in relation to Fallis' account, where it was suggested that agendas typically do not involve both questions about what most people would consider paradigmatically pressing concerns and fairly banal questions about TV shows. That this is so may be granted by Fallis, but the question remains: What if such a span of topics actually happened to be considered most interesting by an inquirer, however atypical she, thereby, would be? By way of illustration, consider a subject who can choose between utilizing two practices,  $P_1$  and  $P_2$ .  $P_1$  would enable her to answer a very small number of questions that figure on her agenda about the causes, treatment, and prevention of cancer.  $P_2$ , on the other hand, would enable her to answer a great many questions, also figuring on her agenda, about a series of utterly banal matters, primarily pertaining to the plots of a couple of soap operas that she follows.<sup>11</sup> Let us also assume that there is no overlap in the questions answered by the two processes, and that the questions answered by  $P_2$  are so banal that no pursuit of challenging questions would make any difference to her ability to answer them. If so, the account defended here implies that  $P_2$  is epistemically better than  $P_1$  for the subject.

Does this amount to a *reductio* of the present account? I do not think that it does. There are a number of ways in which the subject in question would be better off by opting for  $P_1$ . From a moral perspective, the questions answered by  $P_1$  have a potential to

---

<sup>11</sup> Feldman (2001) invokes a similar scenario in a critique of Goldman (1999).

help people suffering from serious and often terminal diseases. From the perspective of personal well-being, it might be that, if the subject is anything like most of us, engaging with the questions addressed by  $P_1$  would provide her with a deeper sense of purpose and intellectual stimulation than engaging with the questions addressed by  $P_2$  does. And so on. However, in developing a notion of epistemic betterness, we are interested in a *specialized* form of evaluation, namely the kind of evaluation that—if the observations made above are on the right track—pertains to the generation of high ratios and numbers of beliefs that answer whatever questions the relevant inquirers happen to be posing. As such, it is to be expected that the present account will not capture aspects relevant to *other* forms of evaluation, such as those just mentioned. Consequently, despite the fact that opting for  $P_2$  makes the subject inferior in all aforementioned respects, she is not, thereby, *epistemically* worse off, and suggesting that she is would simply amount to exaggerating if not misconstruing the scope of epistemic evaluation.

A similar point can be made in relation to considerations about the epistemic evaluation of harmful pursuits. By way of illustration, consider a scenario exactly like the one above, the only difference being that the questions answered by  $P_2$  pertain to how to bring about the maximum amount of gratuitous pain and suffering in the maximal amount of people. For reasons perfectly analogous to those just discussed, the present account would imply that the subject is better off by opting for  $P_2$  rather than for  $P_1$ . Is this a problem? In answering this question, it might be fruitful to consider an account on which this would *not* follow. Recently, Michael Bishop and J. D. Trout (2005) have suggested that “the significance of a problem for  $S$  is a function of the weight of the objective reasons  $S$  has for devoting resources to solving that problem” (p. 95). As for the specific

nature of such objective reasons, they make clear that their “general account of epistemic significance resides, ultimately, in judgments about what conduces to human well-being” (p. 99). Bishop and Trout do not say explicitly whose well-being is relevant to whose objective reasons. However, most of their comments—such as that the “priorities of the excellent reasoner (and more generally, of the wise person) are set so that they may serve as a means to human flourishing” (p. 101)—suggest that the objective weight of *S*’s epistemic reasons is a function of increases in the amount of *general* well-being in the world, rather than simply in the particular well-being of *S*.

Incorporating this notion of significance in the present account of epistemic betterness would imply that the subject would be epistemically better off by opting for  $P_1$  rather than for  $P_2$ . However, there are two problems with forging such a strong connection between significance and well-being. First, what is the normative force of this connection? Is it that we ought to *only* engage in endeavors that are highly conducive to general well-being? If this is the idea, the account seems to not only promote an extremely skewed division of cognitive labor, but also render the majority of what makes up current scientific (not to mention philosophical) research insignificant. In fact, even a more lax construal of conduciveness would be problematic, since there are substantial sets of questions and problems that lack *any* connection to the promotion of well-being. For example, is it always possible to find, somewhere in the infinity of integers, a progression of any length of equally spaced prime numbers? Are there three, four or eleven dimensions to our universe? It is not clear that either of these questions may be related to

matters of human-well being, or that this ought to imply that the corresponding beliefs should not be considered epistemically significant.<sup>12</sup>

That being said, it might be possible to rework Bishop and Trout's account in a way so that the normative force does not flow from an injunction to *maximize* well-being, but from a desire to ensure that inquiry is not *detrimental* to well-being. Still, a fundamental point remains, which brings us to the second problem: While there is, clearly, something to be said for the fact that the subject should not opt for  $P_2$ , it is not obvious that she should be considered *epistemically* worse off by doing so. As noted above, epistemic evaluation is concerned with evaluating practices along a specifically epistemic dimension—a dimension that, I have argued, should be understood in terms of beliefs the contents of which answer whatever questions happen to figure on the inquirer's agenda. As such, we should, as already noted, expect there to be aspects of betterness not captured by the present framework, i.e., *non-epistemic* aspects. However, rather than indicating a flaw in the present account, the fact that it does not account for such aspects simply suggests that it does not account for more than it should, *qua* an account of epistemic betterness.

---

<sup>12</sup> Could the account be saved by appeal to the increase in well-being that may be experienced by the researchers engaged in these ventures (*ceteris paribus*)? This would imply that, if everyone engaging in foundational issues in mathematics, natural science, and metaphysics (among other areas) ceased to experience any degree of well-being in their pursuits, and no one else started doing so for them, the corresponding beliefs would cease to be significant—and for that very reason. I suspect that this would give us a notion of significance very different from what Bishop and Trout are looking for.

### 4.3. Conclusion

The previous two chapters argued that true belief is the only thing of non-instrumental epistemic value. However, it does not follow from true belief being the *only* good of non-instrumental epistemic value that *all* true beliefs are valuable thus. More specifically, the present chapter argued that the epistemic value of a practice is a function of the extent to which it yields beliefs that answer questions posed by relevant sets of inquirers. As such, the account tells us something about what methods or practices inquirers should utilize, in so far as they want answers to their questions. The remainder of the chapter was primarily concerned with tracing the limits of this ‘should,’ and it was suggested that there are very real limits to the extent to which we can make *epistemic* judgments about what questions inquirers should pose, as opposed to about what practices to opt for, *given* a set of questions. However, rather than indicating a flaw in the account, this simply suggests that it is able to account for the specialized nature of epistemic evaluation.

## CHAPTER 5

### EPISTEMIC AGENCY AND PATERNALISM:

#### WHEN LESS IS MORE

##### 5.1. Reasons for Mandating Compliance

When conducting inquiry we do a variety of things, such as gather information, mull over our data, and choose among different methods of investigation—all for the purpose of attaining our epistemic goal. In so doing, we are expressing our *epistemic agency*, as was suggested in chapter 1. Chapters 2 and 3 provided an account of the relevant goal, i.e., the goal of believing truly, and chapter 4 developed a theory about what it is for some practices to be epistemically better than others with respect to that goal, in terms of how good of a job they do in generating beliefs that answer questions posed by the relevant inquirers. The purpose of the present chapter is to illustrate how we may use the resulting notion of epistemic agency to say something constructive about how the ways in which we *actually* express our agency may be brought in line with how we *should* express our agency.

To this end, consider the fact that it, intuitively, might seem that, the greater our freedom to express our epistemic agency, the better. At the very least, it seems *prima facie* plausible that gathering *more* information, spending *more* time mulling over our data, or choosing from a *wider* selection of methods cannot be a bad thing from the point of view of forming true belief with respect to whatever matters happen to be under investigation. However, the present chapter argues that, when it comes to our freedom to

express such agency, more is *not* always better. For example, it will be argued that we often are better off by gathering only a very limited amount of information, having our selection of methods be greatly restricted, and spending our time less on reflecting than on simply reading off a market price or the output of a simple algorithm. In fact, we are in many situations better off from the point of view of attaining true belief by having our agency be constrained in substantial ways, even if against our will.

The case for this claim starts out with two psychological observations. First, as noted already above, it is well known that we suffer from a variety of cognitive biases when reasoning under uncertainty.<sup>1</sup> In the psychological literature, a cognitive bias typically designates a systematic tendency to form judgments in a manner that fails to accord with principles that (given certain normative assumptions) may be gleaned from logic, probability theory, statistics, and related disciplines. Understood thus, certain biases may very well be adaptive, and, as such, yield accurate judgments in many situations, despite being arrived at by way of heuristics that fail to accord with aforementioned principles.<sup>2</sup> In light of this, a more useful way to understand what it is for something to be a bias is in terms of systematic tendencies to form inaccurate belief—be it due to a failure to form belief in accordance with principles gleaned from logic, probability, statistics, etc., or to substantial differences between the cognitive challenges that faced our evolutionary ancestors and the cognitive challenges that face us today.

Second, any attempt at *de-biasing*—i.e., any attempt to either correct or sidestep our systematic tendencies to form inaccurate beliefs through corrective or substitutive strategies, respectively—has to take into account another fact touched upon above,

---

<sup>1</sup> See Gilovich *et al.* (2002).

<sup>2</sup> See, e.g., Gigerenzer *et al.* (1999).

namely that each and every one of us tends to underestimate the extent to which we, as opposed to everyone else, actually are prone to such biases.<sup>3</sup> This is relevant since there are (at least) two ways in which an attempt at de-biasing may fail. On the one hand, it may fail to improve our reliability (within the relevant domain), even if the methods prescribed are utilized properly. On the other hand, the methods prescribed may fail to be utilized at all. Let us refer to the latter possibility as involving *problems of defection*.

Problems of defection have received ample attention within the literature on predictive modeling in connection with the so-called “broken leg problem.”<sup>4</sup> The problem is illustrated with reference to an imagined prediction model (more on such models below) that is highly reliable in predicting a person’s weekly attendance at a movie, but that should be disregarded upon finding out that the person in question has a fractured femur. There is, certainly, something to be said for being sensitive to information not taken into account by whatever models one happens to be relying on in prediction, especially in light of the plausible observation that no domain-specific method will be reliable in all domains. The problem is just that, in a wide range of domains where we rightly care about accuracy of judgment—including the domains of medical and financial decision making—people tend to see far more broken legs than there really are, and, thereby, end up defecting from reliable models far more often than they should, from an epistemic point of view.<sup>5</sup>

Evidence suggests that that major culprit behind defection is that we are systematically overconfident about the accuracy of our judgments. For example,

---

<sup>3</sup> See Pronin *et al.* (2002).

<sup>4</sup> See Meehl (1954).

<sup>5</sup> See Dawes *et al.* (2002).



Whitecotton (1996) found a correlation between confidence ratings and defection from models known to be highly reliable in an earnings-forecasting study with professional financial analysts, as did Arkes *et al.* (1986) in a similar study. The suspicion that this correlation consisted in a relation of causation has been substantiated more recently in Sieck and Arkes (2005). Several studies have suggested that overconfidence is a very recalcitrant phenomenon that is mitigated neither by accuracy incentives<sup>6</sup> nor by simple motivational declarations.<sup>7</sup> What *have* been shown to reduce overconfidence to some degree, however, are rigorous regiments of feedback.<sup>8</sup> Building upon this fact, Sieck and Arkes found that such feedback not only lowered people's overconfidence in their judgments in so far as they did not rely on the statistical models offered, but also led to greater reliance on those models (with a resulting improvement in performance), as compared to a control group. This provides us with reason to believe that overconfidence is one (albeit not necessarily the only) cause of many cases of defection.

This has implications for how we should go about the business of de-biasing. Since defection is the result of such a *widespread* phenomenon as overconfidence, we can expect that a great majority of us will be prone to defection. Moreover, since defection is due to a psychological feature of *us*, as opposed to a feature of the methods typically invoked in attempts at de-biasing, we can also expect that a great majority of us will be prone to defection, no matter the particular characteristics of the methods in question. Granted, it might be suggested that the empirical results in question suggests one way to come to terms with problem of defection, namely by focusing our attention more directly

---

<sup>6</sup> See Arkes *et al.* (1986).

<sup>7</sup> See Lord *et al.* (1984).

<sup>8</sup> See Arkes *et al.* (1987).

on strategies that temper people's confidence in their own accuracy. However, the recalcitrant nature of overconfidence, together with the rigorousness of the feedback schedule required to come to terms with it, renders the practical prospects of mitigating the former by way of the latter somewhat dim.

For this reason, a different strategy will be pursued presently. More specifically, since the above results suggest that a great majority of us will be prone to defection, no matter the particular methods involved, the most promising approach to de-biasing will be one that simply does not provide us with the option of *not* using the methods in question. When the option of not using the relevant methods is removed thus, i.e., when the inquirers, in effect, are left with no option but to utilize the methods in question, we are *mandating compliance* with those methods. More specifically, I submit that the aforementioned research on our dual tendency for bias and overconfidence makes plausible the following:

- (C) For every judgment domain  $D$ , such that we (a) care about accuracy in judgment within  $D$ , and (b) have methods available that (if used properly) have been shown to make those making judgments within  $D$  better off from an epistemic point of view than do the methods involved in prevailing practices, we have *pro tanto* reason for mandating compliance with the former methods within  $D$ .

A couple of comments are in order. First, the question of what it is for a method to make someone better off from an epistemic point of view will be postponed until the next

section, although the discussions in the previous chapter should provide substantial hints as for how this locution will be spelled out. Second, no necessary and sufficient conditions will be given for what constitutes a judgment domain. Instead, subsequent discussions will rely on paradigmatic examples, such as political polling and medical diagnosis and prognosis. While working with paradigmatic examples rather than clear-cut definitions, clearly, will leave room for some degree of ambiguity, that will present no problem presently, as long as it can be agreed that the domains in question contain many examples of judgment tasks that we typically want to get right, no matter the particular ways in which the domains may be further specified. The presence of such tasks is crucial—and, moreover, provides the motivation for including (a) in (C)—since it is not clear that the mere *availability* of epistemically beneficial methods gives us any reason to implement them, unless we care about accuracy within the domain(s) in question. But when we do care thus, and relying on the methods in question have been shown to make us better off from an epistemic point of view as compared to the methods involved in prevailing practices—as per (b) in (C)—we have *pro tanto* reasons to mandate compliance.

The remainder of the chapter is concerned with spelling out the implications of (C). §5.2 argues that the relevant mandate is *epistemically paternalistic*, and briefly discusses a number of domains in which such paternalism is already practiced in society. §5.3 shows that the precedents in question do not imply that a commitment to the relevant kind of paternalism is trivial, by discussing two domains that satisfy (C) but where the mandates in question are still to be imposed, and doing so would be potentially controversial. §5.4 addresses the worry that, if we restrict ourselves to *pro tanto*

reasons—i.e., genuine but not necessarily *decisive* reasons, since the overall balance of reasons may direct us to act differently—it might turn out that we have *non-epistemic* reason not to impose paternalistic constraints on agency, even in cases where the practices in question are epistemically beneficial. In response to this, it is argued that the reasons involved in the cases considered actually amount to reasons all things considered. The paper concludes in §5.5 that there, hence, are several cases in which we have not only *pro tanto* but all things considered reasons for having our epistemic agency be constrained on a wider scale than it currently is.

## 5.2. What is Epistemic Paternalism?

The idea that constraints on our agency can make us better off is not unique to the present investigation. For example, Jon Elster (2000) mounts a persuasive case for agents often being able to improve their lives in a variety of respects by precommitting themselves to certain future actions, either directly or by severely restricting the number and kinds of options that will be available to them at a future time. However, the mandate in (C) should *not* be understood in terms of such self-binding. If it were only for the fact that we suffered from cognitive bias, self-binding might have provided a viable option for de-biasing. However, given our tendency to also underestimate the extent to which we are at all prone to bias, there is simply no reason to believe that we—as in: each and every one of us—will see any reason to bind ourselves to the methods in question (and even if we did, we would subsequently consider ourselves having ample reason to break that commitment, as in aforementioned cases involving defection from prediction models). In

short, due to our dual tendencies for bias and overconfidence, we cannot rely on ourselves for doing what is best for us, from an epistemic point of view.<sup>9</sup>

For this reason, and as indicated above, a different and largely neglected approach to de-biasing will be considered, involving epistemically paternalistic practices. One of the few previous discussions of such practices is that of Goldman (1992b).<sup>10</sup> Goldman suggests that certain forms of information control practiced in society are motivated with reference to how they protect us from our cognitive failings, but he neither discusses the important role our tendencies for overconfidence play in motivating such protection, nor dwells on the issue of defining the relevant kind of epistemic paternalism. Since we discussed the former issue in §5.1, let us consider the latter. Judging by the vast literature on paternalism in social and political philosophy, the conceptual core of paternalism involves *an interference with the liberty of another without her consent but for her own good*.<sup>11</sup> In light of this, we may define what it is for a practice to be *epistemically paternalistic* as follows:

- (P) A practice is epistemically paternalistic if and only if it
  - (i) constrains inquirers' freedom to conduct inquiry in whatever way they see fit;
  - (ii) is implemented without the consent of those constrained; and

---

<sup>9</sup> Elster (2000, ch. 2) does discuss other-binding, but only in terms of cases where the binding in question is not for the benefit of those bound, as in the kind of de-biasing that concerns us here, but simply for the protection of the binder(s).

<sup>10</sup> However, see Sunstein and Thaler (2008) for a recent defense of what they refer to as *libertarian* paternalism, invoked as a tool for ridding ourselves of a variety of behaviors that are systematically detrimental to our well-being. See also Trout (2009 and 2005), who will be discussed further below.

<sup>11</sup> See Young (2008) and Dworkin (2002) for two overviews of relevant discussions.

(iii) is prescribed for the purpose of making those constrained better off from an epistemic point of view.

A couple of comments are at place. (i) is meant to capture the *interference* of paternalistic practices. I will postpone further discussion of this aspect until we are considering real-world cases below. (ii) is meant to capture the *arrogance*, if you will, of paternalistic practices. Note, however, that this condition does not require acting *against* the consent of those constrained. For example, the satisfaction of (ii) is compatible with the inquirer in question appreciating the merits of the relevant methods and happily utilizing them. What is crucial as far as (ii) is concerned is *not* that the inquirer is forced to do something that she does not want to do, but simply that no effort has been made to seek her consent on the issue.<sup>12</sup>

Condition (iii) is meant to capture what makes the relevant form of paternalism *epistemic*.<sup>13</sup> The kind of improvement relevant here can be defined in terms of the notion of epistemic betterness developed in the previous chapter. More specifically, we may think of what it is for someone to be made better off in the relevant sense in terms of how the relevant paternalistic practice outperforms prevailing practices by mandating compliance with methods that make her more reliable, without making her worse off in some other epistemically relevant respect, e.g., with respect to question-answering power. Understanding improvement specifically in terms of such *pareto* improvements in reliability enables us to ignore cases of epistemic incomparability, as discussed in the

---

<sup>12</sup> This is in line with Dworkin (2002).

<sup>13</sup> It is compatible with a practice being epistemically paternalistic that it is motivated by non-epistemic *in addition* to epistemic considerations.

previous chapter. To see why, consider two practices,  $P$  and  $P+$ , where the latter is a *pareto* improved version of the former. If  $P+$  is a *pareto* improved version of  $P$ , then  $P+$  will be better than  $P$  with respect to reliability, and equally good or better with respect to question-answering power. In either scenario,  $P+$  outperforms  $P$ , as defined in the previous chapter.<sup>14</sup>

A further qualification is needed. In the previous chapter, we defined the relevant reliability ratios not in terms of true belief *per se*, but in terms of *significant* true belief, where a belief is significant if and only if it constitutes a sufficiently informative answer to questions on the relevant agent's research agenda. In light of this, we will need to make two simplifying assumptions. First, since significance is a function of the relevant agents' research agenda(s), we will need to assume that those involved in the compared practices are posing the *same questions*, at least as far as the relevant practices are concerned. Second, we also need to assume that all agents involved function normally, in that they would be able to identify sufficiently informative answers as such when they grasp them. To make matters tractable, let us include these factors under a *ceteris paribus* clause, and say that what it is for those constrained to be made better off from an epistemic point of view is for the relevant paternalistic practice to yield a *pareto* improvement with respect to reliability for those constrained, *ceteris paribus*.<sup>15</sup>

If we understand epistemic paternalism along the lines of ( $P$ ), there does seem to be practices already in place in society that would qualify as epistemically paternalistic.

---

<sup>14</sup> The reader may return to the table in §4.1.4 to confirm that this is so.

<sup>15</sup> Why not drop the *pareto* qualification and simply talk about an improvement with respect to reliability, *ceteris paribus*? Because that would not count as improvements in cases in which a practice makes the relevant agents better off with respect to reliability *and* power. Such improvements would constitute *pareto* improvements but not *ceteris paribus* improvements with respect to reliability.

Goldman (1992*b*) discusses four such practices, pertaining to (a) the control of legal evidence on part of U. S. judges (particularly with respect to character evidence), (b) curriculum selection in education, (c) the Federal Trade Commission's actions against false or deceptive advertisement, and (c) deliberate attempts within the news media to simplify matters that are to be presented to the public. In the interest of adding to the literature rather than simply repeating what has already been said, I will not dwell on these practices. It suffices to note that they all serve to constrain peoples' freedom to conduct inquiry in whatever way they see fit by way of restrictions on information access, without seeking their consent, for the purpose of making them better off epistemically by protecting them from their own ignorance or bias in contexts where we care about accuracy. As such, the practices in question are epistemically paternalistic, on the above definition.

Any attempt to elaborate on the need for epistemic paternalism needs to answer to a series of worries, three of which will be considered presently. Consider, first, the point that the motivation for setting up epistemically paternalistic practices, as outlined above, is a set of statistical facts about our cognitive tendencies. As such, there are likely to be individual exceptions, in the sense that there will be some people who do not exhibit the tendencies in question, even if the relevant statistics suggest that they will constitute a minority. In light of this, the worry goes, is it really defensible to be paternalistic not only with respect to the majority that is likely to benefit epistemically, but also with respect to the few subjects who might either not be made better off, or that in some rare instances



perhaps even will be made worse off from an epistemic point of view for being imposed upon thus?<sup>16</sup>

Two lines of response are available here. On the first one, we press the point that the minorities are just that, i.e., *minorities*, and that the few that are not made better off or that are made worse off, most likely, constitute statistical outliers. However, the problem with this line of reasoning is that it seems to presuppose that, *unless* the minority is made up of outliers, the relevant paternalistic practices would not be defensible. That seems to be granting too much, which brings us to the second line of response: In mandating compliance, the goal is not to make each and every one of those constrained better off; it is simply to bring about a positive and hopefully significant epistemic net effect in the relevant population of inquirers. Granted, the smaller the proportion of subjects that do *not* exhibit the dual tendency of bias and overconfidence, the greater the net effect. But it is sufficient for motivating mandated compliance with epistemically virtuous methods that those who *do* exhibit the relevant tendencies are in a majority and we, consequently, can expect there to be such a net effect, irrespective of whether those who do *not* exhibit the relevant tendencies constitute outliers or simply a minority.

In so far as this response is deemed unsatisfactory, there might be a somewhat different worry lurking in the background, pertaining less to the concern for maximizing net epistemic effects among *populations* of inquirers, and more to a regard for the integrity of the individual *inquirers*. This brings us to the second worry, according to which epistemic paternalism is objectionable simply on account of being paternalistic. The closest we come to a statement to this effect in the epistemological literature is in

---

<sup>16</sup> Thanks to Adrian Staub for raising this question.

Larry Laudan's (2006) treatment of evidence control in law. Laudan is explicitly concerned with the epistemic aspects of this practice. Moreover, he describes it as being paternalistic, and clearly intends the term as a pejorative. However, when we turn to his case against the practice in question, we see that Laudan's complaint is that restricting jurors' access to certain kinds of evidence (such as character evidence) "is not a promising recipe for finding out the truth" (p. 25). That is, even *if* Laudan is right in claiming that we have reason to be skeptical about the epistemic merits of prevailing systems of evidence control—and the "if" is important, since Laudan, unfortunately, pays no attention to the body of empirical literature concerned with how jurors process character evidence (e.g., Hunt and Budesheim 2004; Lupfer *et al.* 2000)—his complaint is not that the practice should be given up because it is *paternalistic* but because *it does not work*, as far as enabling jurors to arrive at accurate verdicts is concerned.<sup>17</sup>

Evaluating existing paternalistic practices for their epistemic efficacy is an important task. But it is not one that will be undertaken here, and for the reason just outlined: an argument demonstrating that a particular practice is not fulfilling its intended epistemic purpose is not an argument against *paternalism*; it is an argument for implementing a *different* practice (paternalistic or not). For this reason, §5.4 will return to a more promising way to spell out aforementioned worry, namely in terms of us having *non-epistemic* reason not to impose paternalistic constraints on agency, even in cases where the practices in question are epistemically beneficial. Before considering that possibility, however, the next section will consider a third worry. The worry in question pulls in a completely different direction from the ones just considered, which suggested

---

<sup>17</sup> See also Goldman (1999), who raises concerns that are very similar to Laudan's.

that a commitment to epistemic paternalism is, in a sense, too controversial simply in virtue of being paternalistic. The third and final worry is that the fact that we already practice epistemic paternalism in society, as evidenced by aforementioned examples, indicates that a commitment to epistemic paternalism is not controversial *enough*, in that it carries no substantial implications as for how we ought to set up our epistemic practices.

### 5.3. Two Cases for Expansion

Above, we illustrated the idea of epistemic paternalism with reference to contexts in which we already seem to practice such paternalism in society. This could be an indication to the effect that a commitment to epistemic paternalism is already sufficiently ingrained in the ways in which we set up epistemic practices for any attempt to argue that we should be epistemically paternalistic to provide any reason to do things differently. It is the burden of this section to argue that a commitment to epistemic paternalism is not trivial in this sense. To make this point, it will be argued that a commitment to (C) provides *pro tanto* reason for practicing epistemic paternalism on a wider scale than we are currently doing. This point will be argued with reference to two domains clearly satisfying (C), but where the relevant mandates are still to be imposed and the relevant impositions would be potentially controversial.

### 5.3.1. Prediction Models in Diagnosis and Prognosis

In a recent treatment of the issue of de-biasing, J. D. Trout (2005) considers a variety of “institutional assistance,” as he calls them, for coming to terms with our sometimes quite biased ways. The kinds of assistances he considers include taking advantage of our sensitivity to differences in framing by couching health care information in terms of gains rather than losses in order to encourage preventive medical testing, changing driving behavior by way of chevrons, and allowing employees to direct a portion of their future salary increases toward retirement savings and thereby using our tendencies for inertia and procrastination to our advantage.

Interestingly, however, Trout goes out of his way to argue that, since the measures “assist the person in effectuating their long-term goals or plans, there is no relevant interference, and so [such] regulation is not paternalistic” (p. 409).<sup>18</sup> There is certainly something to be said for the practices not amounting to paternalism, but not because they promote our long-term goals. After all, many paradigmatic instances of paternalistic practices—e.g., the enforcement of laws for wearing a helmet when riding a motorcycle—are motivated exactly on the basis that they promote our long-term goals. What makes the practices Trout considers non-paternalistic is rather that no one is, in any relevantly strong sense, constrained to do (or from doing) anything.

For present purposes, however, the kinds of assistance *not* considered by Trout are more relevant than the ones he does consider. In particular, Trout does not discuss a kind of assistance that he has discussed at length elsewhere (with Michael Bishop in their 2005), and where an illustrative and straightforwardly paternalistic case can be made. The

---

<sup>18</sup> The same approach is taken by Trout more recently in his (2009, p. 210).

practice in question is that of using prediction models as an alternative to relying on the judgments resulting from the unaided reflections of clinicians in medical diagnosis and prognosis. The last fifty years of empirical research on predictive modeling has not only demonstrated the superiority of the former over the latter, but also provided ample evidence of inquirers being very reluctant to comply.<sup>19</sup> Again, this has received some attention within recent epistemological literature, primarily due to Bishop and Trout's discussion. However, what has not been acknowledged is that the literature in question also provides *pro tanto* reason for mandating the use of such models in medical diagnosis and prognosis—or so I will argue.<sup>20</sup>

The first thing to note here is that we typically care about accuracy in medical diagnosis and prognosis, particularly given the potential human costs of inaccurate judgment within these areas. As such, the relevant domains satisfy (C) (*a*). In light of this, it should be considered good news that aforementioned models enable us to increase our reliability significantly in matters of such diagnosis and prognosis, as compared to reliance on unaided clinical judgments. Moreover, there is no reason to believe that such a mandate would, in turn, decrease our question-answering power, i.e., that it would enable us to answer *fewer* questions correctly. Here, it is important that we do not confuse power with what the previous chapter referred to as *robustness* (see §4.1.3). Granted, it is likely that unaided clinical judgment constitutes a more robust method than does reliance on prediction models, since there are many questions for which prediction

---

<sup>19</sup> See Dawes *et al.* (2002).

<sup>20</sup> Despite resisting the term 'paternalism' in his (2005) and (2009), Trout (in correspondence) suggests that there is an implicit case for paternalism in Bishop and Trout (2005). If so, the following can be seen as an attempt to spell out the relevant case explicitly.

models are yet to be developed and, hence, many questions for which the former would outperform the latter. But power, like epistemic betterness, is a property defined for a specific set of questions, i.e., an agenda. In other words, when comparing the power of different methods, we need to hold fixed the questions at issue, such as the questions pertaining to some particular diagnostic or prognostic problem for which a model is available. Given such a set of questions, the claim is that the relevant prediction model will not yield *less* true belief than do the corresponding unaided clinical judgment processes. Consequently, there is good reason to believe that relying on prediction models (whenever available) will lead to a *pareto* improvement in reliability for the inquirers involved, *ceteris paribus*, and that the relevant domains, hence, satisfy (C) (b).

It might be objected that the answers provided by way of prediction models will not be as *informative* as those yielded by unaided clinical reflection. Two things need to be noted here. First, even *if* a case could be made to the effect that unaided clinical judgments generally are more informative than the outputs of prediction models, it is sufficient for the two methods to be on a par on this score that they are both *sufficiently* informative to answer the relevant questions, whether or not one is more informative than the other. Second, it is questionable that a case can be to the effect that unaided clinical judgments generally are more informative than the outputs of prediction models. It certainly cannot be ruled out that unaided clinical reflection often will yield judgments containing *more* information than what can be gleaned from prediction models, which typically yield a simple binary or numerical output. However, even if that is so, the fact that the latter, nevertheless, tend to outperform the former suggests that whatever additional information tends to be provided through unaided clinical judgment generally

is not of a kind that increases accuracy. And when it comes to diagnosis and prognosis, it is accuracy—not linguistic flourishes—that we are ultimately after.

Keeping these points in mind, we see that the judgment domains of medical diagnosis and prognosis, clearly, fall within the scope of (C), which suggests that we have *pro tanto* reason for mandating compliance with the models in question. Moreover, the practice of implementing such a mandate would be epistemically paternalistic. For one thing, the practice would, as already noted, be motivated with reference to how relying on said models makes inquirers epistemically better off relative to prevailing practices (i.e., reliance on unaided clinical judgment). Consequently, the practice would satisfy (P) (iii). Furthermore, a practice that mandated the use of predictive models (when available) would put real restrictions on clinicians' ability to pursue diagnoses and prognoses in whatever manner they happen to see fit. Consequently, the practice would satisfy (P) (i). As noted above, this is not to rule out the possibility that the relevant inquirers may happily utilize the models, perhaps on account of appreciating their merits. However, considering the well-documented prevalence of defection, and the fact that inquirers have been shown to disapprove (even while lacking epistemic reason for doing so), the practice of implementing the mandate can be motivated without any concern for whether the inquirers themselves actually consent to its implementation. Consequently, the practice would also satisfy (P) (ii).

In conclusion, we have *pro tanto* reason to mandate compliance in utilizing those prediction models that have been shown to make us epistemically better off, as compared to prevailing practices. Moreover, the practice of mandating such compliance would be epistemically paternalistic, in virtue of not only being motivated with reference to how it

would make clinician's epistemically better off but also constraining their freedom to conduct inquiry in whatever way they see fit, without seeking their consent in so doing. Given that such a mandate is yet to be instituted, a commitment to (C), hence, provides *pro tanto* reason to practice epistemic paternalism on a wider scale than we are currently doing.

### 5.3.2. Prediction Markets as an Alternative to Social Deliberation

The second domain in which we have *pro tanto* reason to implement paternalistic practices can be illustrated with reference to the fact that, in any reasonably complex society or organization, it will often be the case that only a minority of the people involved will be informed on any given issue. In those domains where we care about accuracy, how are we to mine such informed minorities for information? One intuitively plausible and historically prominent way to do so is by way of social deliberation, understood as the process of a group of people exchanging reasons for and against a (set of) conclusion(s) for the purpose of reaching a social verdict on the issue in question. Does such a process do a good job of eliciting information from informed minorities?

To answer this question, we need to consider what happens to informed minorities in contexts of social deliberation. There are two possibilities. First, assume that the members of the informed minority do *not* disclose their information. After all, there are many circumstances under which people refrain from disclosing what they know (or what they take themselves to know), either in light of the *informational* pressure coming out of whatever happens to be the majority position (and the assumption on part of the minority that they, not the majority, are probably mistaken), or the *social* pressure associated with



the risk of social sanctions against dissenters.<sup>21</sup> In other words, in the event that there is a diversity of opinions, and the informed members find themselves in a minority, there is a real risk that the information they have in their possession will not even be submitted for deliberation.

Second, assume that the members of the informed minority *do* disclose their information. What is likely to be the impact of that information on the deliberating group? Given that the informed members make up a minority, the impact is likely not to be particularly great, due to what is typically referred to as the *common knowledge effect*. (Since the effect in question applies to information submitted for deliberation in general, and not just to information believed in a manner that amounts to knowledge, a more appropriate designation might have been “the common *information* effect”.) More specifically, social psychologists Daniel Gigone and Reid Hastie sum up the relevant research in terms of how “[t]he influence of a particular item of information [on the judgment of a group] is directly and positively related to the number of group members who have knowledge of that item before the group discussion and judgment” (Gigone and Hastie 1993, p. 960).

In other words, what makes a difference when it comes to having an impact on group judgments is not so much *quality* of information as *quantity* of people bringing a particular piece of information to the table. This would be great news, were there a robust correlation between the information had by the majority and the truth. While such a correlation might hold with respect to certain easy questions that have demonstrable answers, it clearly does not hold with respect to a great many of the questions that we

---

<sup>21</sup> See Sunstein (2006) for a discussion of the relevant evidence.

typically deliberate over, such as questions about politics, public policy, public health, scientific research, and so on. Consequently, it does not seem that social deliberation provides a particularly promising method for eliciting and aggregating dispersed pieces of information for the purpose of arriving at informed judgments on a wide variety of questions that we care about answering. At best, it seems a *prima facie* promising method for aggregating information that is already widely shared—which is not the kind of role we are concerned with presently.

This gives us reason to look elsewhere for a more promising method for eliciting information from informed minorities. One strategy that has received a lot of attention recently is that of setting up a prediction market. A prediction market is a market that is designed and run for the primary purpose of (a) mining and aggregating information scattered among traders, and subsequently (b) using this information in the form of market values to make predictions about future events. As it turns out, the market prices of such markets provide surprisingly accurate information about the events bet on.<sup>22</sup>

Moreover, there is evidence to the effect that accuracy of the market prices that arise in prediction market are to a great extent due to small subsets of traders, typically referred to as *marginal traders*.<sup>23</sup> Beyond investing more money in the markets and spending more time trading than the average trader, marginal traders also earn significantly higher returns. While it might be thought that this is a mere artifact of marginal traders simply investing more money than non-marginal ones, further considerations suggest that the most plausible interpretation of these returns is that the

---

<sup>22</sup> See Bragues (2009), Wolfers and Zitzewitz (2006), and Berg *et al.* (2003) on political prediction markets, Pennock *et al.* (2001) on commercial forecasting, Chen and Plott (2002) on sales forecasting, and Servan-Schreiber *et al.* (2004) on sports forecasting.

<sup>23</sup> See Forsythe *et al.* (1999).

former simply are more informed and, consequently, place better bets than the latter. After all, by virtue of their high trading frequency and higher than average investments, marginal traders have a disproportionately high influence on the price signal. Given that the price signals of prediction markets tend to amount to good predictions, this suggests that the accuracy of the predictions of such markets are, to a great extent, due exactly to the informed investments of marginal traders. In fact, there is further and more direct evidence for this claim. Separating out the subset of marginal traders that are particularly active in setting bid and ask prices, as compared to those traders (marginal or not) who mostly accept others' prices, Oliven and Rietz (2004) found that the average error rate for the latter on the 1992 vote share market of the Iowa Electronic Markets—one of the oldest and most prominent electronic prediction markets—was almost six times that of the former (8% across all trades, versus 47%).

The role played by such marginal traders in prediction markets has implications for how to think about the respective merits of social deliberation and prediction markets. More specifically, above discussion suggests that *prediction markets do a good job of mining information exactly under the conditions where social deliberation does not, i.e., in contexts of informed minorities*. In fact, the evidence just reviewed gives us *pro tanto* reason to institute a paternalistic mandate for using information gleaned from prediction markets rather than that coming out of social deliberation, in cases where informed minorities are present. To see why, consider, first, the fact that many of the issues that we care to deliberate about socially, including issues pertaining to governmental policy and organizational strategy, are ones where we care about accuracy. Consequently, the domains in question satisfy (C) (a). Next, consider the relative epistemic merits of relying

on prediction markets as opposed to on social deliberation. More specifically, consider the following two processes:

- (1) Forming beliefs on the basis of the judgments of deliberating groups; and
- (2) forming beliefs on the basis of the price signals of prediction markets.

Given what we know about the workings of social deliberation and prediction markets, respectively, we have good reason to believe that, if informed minorities are present, someone forming belief in the manner described in 2 will form a higher ratio of true belief than someone forming belief in the manner described in 1 (*ceteris paribus*).

Furthermore, in analogy with what was argued in relation to the power of prediction models above (and keeping in mind the distinction between power and robustness), there is no reason to believe that basing ones beliefs on the price signals of prediction markets, whenever available, will lead to a decrease in question-answering power.

Consider, also, the issue of whether the answers gleaned from prediction markets are sufficiently informative. While the social verdicts of deliberating groups may, certainly, provide *more* information than what can be gleaned from the price signals of a prediction market, the fact that the someone relying on the latter as opposed to on the former is likely to be epistemically better off suggests that prediction markets do a significantly better job than do deliberating groups in harnessing the *right* information and, in turn, making for accurate answers. As such, the relevant worry about informativeness is not that prediction markets provide insufficiently informative outputs; it is that socially deliberating groups might not satisfy a necessary condition on any

sufficiently informative answer, as defined in the previous chapter, i.e., that the judgment in question be an accurate one.

In other words, there is good reason to believe that relying on prediction markets (whenever available) will lead to a *pareto* improvement with respect to reliability for the inquirers involved, *ceteris paribus*, and that the relevant domains, hence, satisfy (C) (b). As such, the many domains typically associated with social deliberation fall within the scope of (C), suggesting that we have *pro tanto* reason for mandating reliance on the outputs of relevant prediction markets, whenever available. More than that, a practice involving such a mandate would be epistemically paternalistic. As already noted, the mandate would be motivated with reference to the epistemic merits of relying on prediction markets. As such, it would satisfy (P) (iii). However, it should be stressed that the practice of enforcing the mandate would *not* involve a ban on social deliberation. Instead, it would involve a mandate on those *making judgments* within the relevant domains to use information gleaned from prediction markets (when such markets are available) rather than information produced by way of social deliberation. Since this would put restrictions on the freedom of those making judgments to use whatever kind of information they deem relevant, the practice would satisfy (P) (i). Moreover, since there is far from a guarantee that those restricted thus would welcome the restriction, defection is a real possibility, not the least given the risk that inquirers will exhibit overconfidence in their assessments of whether they will be able to judge what kind of information is most probative. As above, this is not to deny that some of those restricted may actually welcome the restriction; it is simply to suggest that the practice can be motivated without

any concern for whether the inquirers themselves actually approve of its implementation. As such, the practice satisfies (P) (ii).

In conclusion, many of the judgment domains typically associated with social deliberation satisfy (C), and we, consequently, have *pro tanto* reason to mandate reliance on prediction markets as opposed to on the outputs of social deliberation within the relevant domains, whenever such markets are available. More than that, the practice of mandating such reliance would be epistemically paternalistic, in virtue of not only being motivated with reference to how it would make those making judgments within the relevant domains better off, but also constrain their freedom to conduct inquiry in whatever way they see fit, without seeking their prior consent. Given that such a mandate is yet to be instituted on the scale considered presently, we thereby have yet another example of how a commitment to (C) provides *pro tanto* reason to practice epistemic paternalism on a wider scale than we are currently doing.

#### 5.4. Countervailing Reasons

The previous section argued that we have *pro tanto* reason for practicing epistemic paternalism on a wider scale than we are currently doing, and illustrated this point with reference to the use of prediction models in medical diagnosis and prognosis, and the reliance on prediction markets as a tool for mining informed minorities for information. The present section returns to a version of the worry discussed in §5.2, to the effect that paternalistic practices are objectionable simply on account of us having *non-epistemic*

reasons not to impose paternalistic constraints on agency, even if such constraints turn out to be epistemically efficacious.

In order to address this worry, we need to say something about the conditions under which *pro tanto* reasons amount to reasons all things considered. The following is a sufficient condition:

- (S) A *pro tanto* reason to  $\varphi$  amounts to a reason to  $\varphi$  all things considered if there are no *countervailing* reasons, i.e., there are no reasons *not* to  $\varphi$  that are strong enough to offset the *pro tanto* reason to  $\varphi$ .

The final point to be argued in this chapter is that the reasons pertaining to the paternalistic practices discussed in §5.3 amount to reasons all things considered, when the relevant mandates are properly qualified. Let us consider the two cases in turn.

#### 5.4.1. Reasons for Not Relying on Prediction Models

Opting for the output of a statistical model rather than relying on the judgment of experienced clinicians might seem to amount to stripping the relevant patients of their individuality. At least that is how it feels. But does that feeling give us an ethical reason against mandating the use of prediction models? It is not clear that it does, considering the ethical costs of relying on methods that consistently have been shown to be equally reliable at best, and significantly less reliable at worst. In this respect, little has changed since Robyn Dawes wrote the following in a classical review article:

No matter how much we would like to see this or that aspect of one or another of the studies reviewed in this article changed, [...] no matter how ethically uncomfortable we may feel at 'reducing people to mere numbers,' the fact remains that our clients are people who deserve to be treated in the best manner possible. If that means—as it appears at present—that selection, diagnosis, and prognosis should be based on nothing more than the addition of a few numbers representing values on important attributes, so be it. To do otherwise is cheating the people that we serve (Dawes 1979, p. 581).

In other words, it seems that ethical reasons might even speak *for* a strict reliance on prediction models, given the potential ethical costs of relying on a method that has been shown to be equally reliable at best, and significantly less reliable at worst.

Another potential worry is that relying on prediction models will be more costly and time-consuming than relying on unaided clinical judgments. However, the typical prediction model tends to be developed exactly for the purpose of decreasing cost and reducing the time needed to arrive at a diagnosis or prognosis. By way of a representative example, consider the model developed by Breiman *et al.* (1993) to classify potential heart attack patients according to risk, in order to reduce the number of unnecessary admittances of low risk patients to coronary care units. Beyond being more reliable than both unaided clinicians and several complex statistical classification methods, the model only requires the clinician to answer a maximum of three diagnostic, yes/no questions.

By way of a further example, consider the Ottawa Ankle Rule, developed by Stiell *et al.* (1992) as a tool for diagnosing ankle fractures for the purpose of reducing the number of unnecessary x-rays. In using the rule, the clinician needs to answer a maximum of three diagnostic questions in order to determine whether an x-ray is needed. Despite its simplicity, the rule has a very low rate of false negatives, yet allows for a reduction of up to 36% of ankle x-rays, compared to prevailing practices. Still, Graham *et*



*al.* (2001) reports that, while 96% of emergency physicians in the United States are familiar with the rule, 67% never use it, or only use it sometimes.

In other words, considerations about cost and time-efficiency might just be speaking *for* a strict reliance on prediction models, whenever available. This, of course, is compatible with there being high costs associated with *introducing* (as opposed to using) the models in question.<sup>24</sup> This brings us to the third worry: There might be cases in which introducing the model will exceed the benefits of using them. Needless to say, this possibility cannot be ruled out *a priori*, even if the benefits of increased accuracy and decreased cost and time used that result from opting for prediction models make it likely that even a significant start-up cost can be offset over time. Still, it seems reasonable to require that the mandate should not apply in cases where the costs of introducing the models in question clearly exceed the benefits of using them. The relevant weighing of costs and benefit will, most likely, have to be done on a case-by-case basis, in a manner that is, hopefully, sufficiently sensitive to the variety of concerns that may be brought to bear on the issues at hand to, at the very least, reveal cases in which even aggregate benefits are clearly outweighed by relevant costs.

In conclusion, the overall balance of epistemic, ethical, and practical considerations suggest that we have reason to institute a paternalistic mandate on using prediction models as opposed to relying on unaided clinical judgments, unless the cost of introducing the models in question clearly outweighs even the long-term benefits—epistemic and otherwise—of doing so. Given this qualification, and in the absence of any

---

<sup>24</sup> See Bishop and Trout (2005) for a good discussion.

further countervailing reasons, (S) implies that the *pro tanto* reason involved amounts to a reason all things considered.

#### 5.4.2. Reasons for Not Relying on Prediction Markets

Turning now to prediction markets, three concerns deserve mention. First, prediction markets require a fairly demanding infrastructure, and there might be many situations where constraints on cost and other available means (not the least with respect to attracting a sufficiently informed subset of traders) will make setting up a reliable prediction market practically impossible. However, the mandate that concerns us here is not one on *setting up* prediction markets, but rather on *using* the information coming out of such markets, when available. The epistemic merits of using such information might also give us *pro tanto* reason to set up *more* such markets. That, however, is a different matter from the one that we have been concerned with here.

Second, social deliberation is often taken to be what renders governmental policies politically legitimate, particularly in so far as the policies in question involve the use of force against those representing a minority opinion. In light of this, there might be worries about whether a greater reliance on prediction markets as opposed to on social deliberation on part of governments—and perhaps on part of social institutions more generally—would serve to erode the legitimacy of the policies in question.<sup>25</sup>

A couple of things need to be noted in response to this worry. The idea that deliberation is the source of legitimacy is particularly prominent within theories of deliberative democracy, in turn “rooted in the intuitive ideal of a democratic association

---

<sup>25</sup> Thanks to Robert Talisse for raising this issue.

in which the justification of the terms and conditions of association proceeds through public argument and reasoning among equal citizens,” to quote Joshua Cohen (1997, p. 72). The fact that the ideal is taken to apply to the justification of *the terms and conditions of association* is important, since it suggests that the ideal is not meant to apply to all instances of policy making, but only to those policies that settle fundamental issues about what terms and conditions of democratic association would be just or fair. On this interpretation, the deliberative ideal applies only to what John Rawls (1997) in his account of public reason refers to as “constitutional essentials and matters of basic justice” (p. 95), pertaining to issues such as “who has the right to vote, or what religions are to be tolerated, or who is to be assured fair equality, or to hold property” (p. 94). Needless to say, these are *not* the kinds of issues that we are imagining that the price signals of prediction markets are to settle. Rather, prediction markets would enter at the point where questions about constitutional essentials and matters of basic justice are settled, and the relevant questions about governance are predictive questions about what policies, as a matter of empirical fact, are likely to generate what goods, as defined by whatever turns out to be the proper account of justice.

That point aside, imagine that the ideal in fact *does* apply more generally, in that it applies to any procedure invoked in arriving at policies that involve the use of force against minorities. In this context, it is crucial to note that what we are dealing with here is an *ideal*, that, as such, leaves open how we can best approximate this deliberative conception of democratic interaction in practice. As Cohen puts the point, “[t]he appropriate ordering of deliberative institutions depends on issues of political psychology and political behavior; it is not an immediate consequence of the deliberative ideal” (p.

85). Consequently, it cannot be ruled out that a proper *realization* of the deliberative ideal might look very different from the ideal itself. To name but one respect in which this will be so, consider Jürgen Habermas' (1975) conception of ideal deliberation as involving a context in which "no force except that of the better argument is exercised" (p. 108). As we have seen, actual social deliberation is very different from such ideal deliberation in that *several* forces are exercised beyond that of the better argument, including the force exerted by the sheer the number of proponents that a particular argument happens to have, or the social or informational pressures that may suppress the arguments of minorities from even being submitted for deliberation. As such, social deliberation is under many circumstances simply not conducive to the inclusion of all participants' perspectives, nor consequently to the treatment of all deliberating citizens as *equal*, contrary to the picture of democratic deliberation associated with aforementioned intuitive ideal.<sup>26</sup> In particular, and for reasons discussed above, the epistemic perspectives of minorities are bound to lose out in contexts of social deliberation simply on account of being minorities.

As we have seen, prediction markets do a much better job than social deliberation in taking into account the reasons and perspectives provided by minorities, at least in so far as they are informed on the relevant issues. Indeed, in contexts involving such minorities, it seems that setting up a prediction market will actually provide a better approximation of the deliberative ideal than social deliberation would do. This contention might seem paradoxical, but should come as no surprise given (*a*) the idealized nature of

---

<sup>26</sup> This is not to suggest that equality would be sufficient (as opposed to merely necessary) for realizing the relevant ideal. If that were the case, a simple coin toss would do the trick, as pointed out by Estlund (2008).

the intuition driving deliberative conceptions of democracy, and (b) the tendency for bias and overconfidence that we are subject to as actual deliberators. Moreover, keeping this in mind, it is not clear why the minority should feel any *less* compelled to accept decisions based on information gleaned from prediction markets as opposed to from social deliberation, contrary to the worry raised above. In the event that the members of the minority are uninformed, they will run the risk of being ignored irrespective of whether they are trading on a prediction market or attempting to partake in deliberation (in the former case, by virtue of being uninformed; in the latter case, simply by virtue of being members of a minority). More than that, if the minority happens to be informed, its members might even have *more* reason to feel compelled thus in light of decisions taken on the basis of markets rather than on deliberation, since they have a greater chance of having their voices be heard in the former than in the latter.

Third, if there is a mandate on using information from prediction markets, and more decisions, as a result, will be based on the price signals of such markets, we can expect attempts to influence those decisions through market price manipulations. How great of a threat is this? There is plenty of evidence of people investing in political prediction markets with the intention of boosting the perceived likelihood of their candidate winning the relevant office. However, as pointed out by Rhode and Strumpf (2009), there is little evidence that political prediction markets can be systematically manipulated beyond short periods of time. By way of example, political betting markets on Wall Street in the late 19<sup>th</sup> and early 20<sup>th</sup> century involved millions of dollars, and many attempts at boosting candidates by way of speculative investments. These

speculations resulted in price change, but the prices returned to their pre-attack levels within days.

More recently, a series of random investments made by Rhode and Strumpf themselves in the 2000 Iowa Electronic Markets' presidential market led to large initial price changes, but the prices reverted to their initial levels in a few hours. These are, of course, exactly the results we should expect to get in a market where at least some of the investors utilize accurate information for the purpose of profiting from the influx of liquidity provided by the manipulators. That said, while available empirical evidence indicates that manipulation may not be as urgent a problem as it might seem at first glance, the possibility of such manipulation suggests that the mandate to rely on prediction markets should be suspended in the presence of evidence for large-scale strategic investments.

In conclusion, the overall balance of epistemic, practical, political, and ethical considerations come out in favor of a paternalistic mandate on using information gleaned from prediction markets rather than information coming out of social deliberation, in cases where we (a) have reason to believe that informed minorities are present, and (b) lack any evidence for large-scale strategic investments in the relevant markets. Given this qualification, and in the absence of any further countervailing reasons, (S) implies that the *pro tanto* reason involved amounts to a reason all things considered.

## 5.5. Conclusion

When conducting inquiry, we gather information, mull over our data, choose among different methods of investigation, and so on—all for the purpose of attaining true belief relevant to the questions that we pose. In so doing, we are expressing our epistemic agency. Above, it was argued that, when it comes to our freedom to express such agency, more is not always better. These points were argued with reference to how our dual tendency for bias and overconfidence gives us *pro tanto* reason for mandating compliance with methods that have been shown to lead to pareto improvements in reliability. Moreover, it was argued that mandating compliance thus would be *epistemically paternalistic*, and that we have not just *pro tanto* but all-things-considered reason to practice such paternalism on a wider scale than we are currently doing. For example, in many situations, we are better off by gathering only a very limited amount of information, having our selection of methods be greatly restricted, and spending our time less on reflecting than on simply reading off a market price or the output of a simple prediction model. Consequently, when it comes to our freedom to express our epistemic agency, more is not always better. In fact, less is often so much more.

## BIBLIOGRAPHY

- Alston, W., (2005), *Beyond "Justification": Dimensions of Epistemic Evaluation*, Ithaca: Cornell University Press.
- Arkes, H. R., Dawes, R. M., and Christensen, C., (1986), "Factors influencing the use of a decision rule in a probabilistic task," *Organizational Behavior and Human Decision Processes* 37, pp. 93–110.
- Arkes, H., Christensen, C., Lai, C., and Blumer, C., (1987), "Two Methods of Reducing Overconfidence", *Organizational Behavior and Human Decision Processes* 39, pp. 133–144.
- Baehr, J., (2009), "Is There a Value Problem?" in A. Haddock, A. Millar, and D. Pritchard (2009), pp. 42-59.
- Belnap, N., and Steel, T., (1976), *The Logic of Questions*, New Haven: Yale University Press.
- Berg, J., Nelson, F., and Rietz, R., (2003), "Accuracy and Forecast Standard Error of Prediction Markets," manuscript, July 2003.
- Bishop, M., (2000), "In Praise of Epistemic Irresponsibility: How Lazy and Ignorant Can You Be?," *Synthese* 122, pp. 179-208.
- Bishop, M., and Trout, J. D., (2005), *Epistemology and the Psychology of Human Judgment*, New York: Oxford University Press.
- Bohman, J., and Rehg, W., (1997), *Deliberative Democracy: Essays on Reason and Politics*, Cambridge: The MIT Press.
- BonJour, L., (1985), *The Structure of Empirical Knowledge*, Cambridge: Harvard University Press.
- Bradley, B., (2006), "Two Concepts of Intrinsic Value," *Ethical Theory and Moral Practice* 9, pp. 111-130.
- Bragues, G., (2009), "Prediction Markets: The Practical and Normative Possibilities for the Social Production of Knowledge," *Episteme* 6 (2), pp. 91-106.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., (1993), *Classification and Regression Trees*, Boca Raton: Chapman & Hall.



- Burnyeat, M. F., (1980), "Socrates and the Jury: Paradoxes in Plato's Distinction Between Knowledge and True Belief," *Proceedings of the Aristotelian Society* suppl. vol. 54, pp. 173-191.
- Chang, R., (1997), ed., *Incommensurability, Incomparability, and Practical Reason*, Cambridge: Harvard University Press.
- Chang, R., (2002), "On the Possibility of Parity," *Ethics* 112, pp. 659-688.
- Chen, K., and Plott, C., (2002), "Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem," CalTech Social Science Working Paper No. 1131.
- Chisholm, R. M., (1968), "Lewis' Ethics of Belief," in P. Schilpp, ed., *The Philosophy of C. I. Lewis*, LaSalle: Open Court.
- Chisholm, R. M., (1977), *Theory of Knowledge*, 2<sup>nd</sup> ed., Englewood Cliffs: Prentice-Hall.
- Chisholm, R. M., (1991), "Firth and the Ethics of Belief," *Philosophy and Phenomenological Research* 51 (1), pp. 119-128.
- Clifford, W. K., (1866), "The Ethics of Belief," in L. Stephen and F. Pollock, ed., *Lectures and Essays by the late William Kingdon Clifford*, 2<sup>nd</sup> ed., London: Macmillan and Co., pp. 339-363.
- Cohen, J., (1997), "Deliberation and Democratic Legitimacy," in Bohman and Rehg (1997), pp. 67-91.
- Conee, E., and Feldman, R., (2008), "Evidence," in Q. Smith, ed., *Epistemology: New Essays*, New York: Oxford University Press, pp. 83-104.
- Cooper, J. M., (1997), *Plato: Complete Works*, Indianapolis: Hackett.
- David, M., (2001), "Truth as the Epistemic Goal", in M. Steup, ed., *Knowledge, Truth, and Duty*, New York: Oxford University Press.
- David, M., (2005), "Truth as the Primary Epistemic Goal: A Working Hypothesis," in M. Steup and E. Sosa, (2005), pp. 296-312.
- Dawes, R., (1979), "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist* 34 (7), pp. 571-582.
- Dawes, R., Faust, D., Meehl, P., (2002), "Clinical versus Actuarial Judgment," in Gilovich et al. (2002), pp. 716-729.

- DePaul, M. R., (2009), "Ugly Analyses and Value," in A. Haddock, A. Millar, and D. Pritchard (2009), pp. 112-138.
- Descartes, R., (1641/1984), "Meditations on First Philosophy," in J. Cottingham, R. Stoothoff, and D. Murdoch, eds., *The Philosophical Writings of Descartes vol. II*, Cambridge: Cambridge University Press.
- Dworkin, G. (2002) "Paternalism," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*; accessed from <http://plato.stanford.edu/entries/paternalism/> in November, 2009.
- Elster, J., (2000), *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*, Cambridge: Cambridge University Press.
- Estlund, D. M., (2008), *Democratic Authority: A Philosophical Framework*, Princeton: Princeton University Press.
- Fallis, D., (2006), "Epistemic Value Theory and Social Epistemology," *Episteme* 2 (3), pp. 177-188.
- Feldman, R., (1985), "Reliability and Justification," *the Monist* 68, pp. 159-174.
- Feldman, R., (1988), "Epistemic Obligations," *Philosophical Perspectives* (2), pp. 235-356.
- Feldman, R., (1999), "Methodological Naturalism in Epistemology," in J. Greco and E. Sosa, eds., *The Blackwell Guide to Epistemology*, Malden: Blackwell, pp. 170-186.
- Feldman, R., (2001), "Review of Alvin Goldman's Knowledge in a Social World," *British Journal for the Philosophy of Science* 52, pp. 163-168.
- Feldman, R., (2002), "Epistemological Duties," in P. Moser, ed., *The Oxford Handbook of Epistemology*, New York: Oxford University Press, pp. 361-384.
- Feldman, R., (2003), "Chisholm's Internalism and its Consequences," *Metaphilosophy* 34 (5), pp. 603-620.
- Field, H., (2001), *Truth in the Absence of Fact*, New York: Oxford University Press.
- Fine, G., (2003), *Plato on Knowledge and Forms: Selected Essays*, New York: Oxford University Press.
- Fine, G., (2004), "Knowledge and True Belief in the *Meno*," *Oxford Studies in Ancient Philosophy* XXVII, pp. 41-81.

- Fine, G., (2008), *The Oxford Handbook of Plato*, New York: Oxford University Press.
- Forsythe, R., Rietz, T. A., and Ross, T. W., (1999), "Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets," *Journal of Economic Behavior and Organization* 39, pp. 83-110.
- Gigerenzer, G., Todd, P. M., and the ABC Research Group, eds., (1999), *Simple Heuristics that Make Us Smart*, New York: Oxford University Press.
- Gigone, D., and Hastie, R., (1993), "The Common Knowledge Effect: Information Sharing and Group Judgment," *Journal of Personality and Social Psychology* 65 (5), pp. 959-974.
- Gilovich, T., Griffin, D., and Kahneman, D., (2002), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press.
- Goldman, A., (1978), "Epistemics: The Regulative Theory of Cognition," *The Journal of Philosophy* 75 (10), pp. 509-523.
- Goldman, A., (1986), *Epistemology and Cognition*, Cambridge: Harvard University Press.
- Goldman, A., (1992), *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge: The MIT Press.
- Goldman, A., (1992a), "Foundations of Social Epistemics," in Goldman (1992), pp. 179-208.
- Goldman, A., (1992b), "Epistemic Paternalism: Communication Control in Law and Society," in Goldman (1992), pp. 209-226.
- Goldman, A., (1999), *Knowledge in a Social World*, New York: Oxford University Press.
- Goldman, A., (forthcoming), "Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism's Problems, Reliabilism's Rescue Package," in T. Dougherty, ed., *Evidentialism and Its Discontents*, New York: Oxford University Press.
- Goldman, A., and Olsson, E. J., (2009), "Reliabilism and the Value of Knowledge," in A. Haddock, A. Millar, and D. Pritchard (2009), pp. 19-41.
- Graham, I. D., Stiell, I. G., Laupacis, A., McAuley, L., Howell, M., Clancy, M., Durieux, P., Simon, N., Empanaza, J. I., Aginaga, J. R., O'Connor, A., and Wells, G., (2001), "Awareness and use of the Ottawa Ankle and Knee Rules in 5 countries: Can Publication Alone be Enough to Change Practice?" *Annals of Emergency Medicine*, 37(3), pp. 259-266.

- Greco, J., (2009), "The Value Problem," in A. Haddock, A. Millar, and D. Pritchard (2009), pp. 313-321.
- Haack, S., (1993), *Evidence and Inquiry: Towards Reconstruction in Epistemology*, Oxford: Basil Blackwell.
- Haack, S., (2001), "'The Ethics of Belief' Reconsidered," in M. Steup, ed., *Knowledge, Truth, and Duty*, New York: Oxford University Press, pp. 21-33.
- Habermas, J., (1975), *Legitimation Crisis*, translated by T. McCarthy, Boston: Beacon Press.
- Haddock, A., Millar, A., and Pritchard, D., eds., (2009), *Epistemic Value*, New York: Oxford University Press.
- Heil, J., (1984), "Doxastic Incontinence," *Mind* 93 (369), pp. 56-70.
- Hempel, C., (1965), "Aspects of Scientific Explanation," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press, pp. 331-496.
- Hunt, J. S., and Budesheim, T. L., (2004), "How Jurors Use and Misuse Character Evidence," *Journal of Applied Psychology* 89 (2), pp. 347-361.
- James, W., (1907), "The Will to Believe," in *The Will to Believe and Other Essays in Popular Philosophy*, New York: Longmans Green and Co., pp. 1-31.
- Jaworski, W., (2009), "The Logic of How-Questions," *Synthese* 166, pp. 133-155.
- Jones, W. E., (1997), "Why Do We Value Knowledge?" *American Philosophical Quarterly* 34 (4), pp. 423-439.
- Kitcher, P., (1992), "The Naturalists Return," *The Philosophical Review* 101 (1), pp. 53-114.
- Klein, P., (1981), *Certainty*, Minneapolis: University of Minnesota Press.
- Kornblith, H., (1994), *Naturalizing Epistemology*, 2<sup>nd</sup> ed., Cambridge: The MIT Press.
- Kornblith, H., (2002), *Knowledge and Its Place in Nature*, New York: Oxford University Press.
- Kornblith, H., (forthcoming), *On Reflection*, New York: Oxford University Press.
- Korsgaard, C., (1996), *The Sources of Normativity*, Cambridge: Cambridge University Press.

- Kvanvig, J., (2003), *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.
- Kvanvig, J., (2009), "The Value of Understanding," in A. Haddock, A. Millar, and D. Pritchard (2009), pp. 95-111.
- Latus, A., (2000), "Our Epistemic Goal," *Australasian Journal of Philosophy*, 78 (1), pp. 28-39.
- Laudan, L., (1977), *Progress and Its Problems: Towards a Theory of Scientific Growth*, Berkeley: University of California Press.
- Laudan, L., (2006), *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*, Cambridge: Cambridge University Press.
- Locke, J., (1706/1996), "Of the Conduct of the Understanding," in R. W. Grant and N. Tarcov, eds., *Some Thoughts Concerning Education and Of the Conduct of the Understanding*, Indianapolis: Hackett.
- Lord, C. G., Preston, E., and Lepper, M. R., (1984), "Considering the Opposite: A Corrective Strategy for Social Judgment," *Journal of Personality and Social Psychology* 47 (6), pp. 1231-1243.
- Lupfer, M. B., Weeks, M, and Dupuis, S., (2000), "How pervasive is the negativity bias in judgments based on character appraisal?" *Personality and Social Psychology Bulletin* 26, pp. 1353–1366.
- Lynch, M., (2004), *True to Life: Why Truth Matters*, Cambridge: The MIT Press.
- Lynch, M., (2009a), "The Values of Truth and the Truth of Values," in A. Haddock, A. Millar, and D. Pritchard (2009), pp. 225-242.
- Lynch, M., (2009b), "Truth, Value and Epistemic Expressivism," *Philosophy and Phenomenological Research* 79 (1), pp. 76-97.
- Meehl, P. (1954), *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Minneapolis: University of Minnesota Press.
- Mill, J. S., (1861/2001), "Utilitarianism"; reprinted in G. Sher, ed., *Utilitarianism and the 1868 Speech on Capital Punishment*, 2<sup>nd</sup> ed., Indianapolis: Hackett Publishing Company, 2001, pp. 1-64.
- Moran, R., (2001), *Authority and Estrangement*, Princeton: Princeton University Press.

- Oliven, K., and Rietz, T. A., (2004), "Suckers are Born but Markets are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market," *Management Science* 50 (3), pp. 336-351.
- Olsson, E. J., (2005), "The Impossibility of Coherence," *Erkenntnis* 63, pp. 387-412.
- Olsson, E. J., (2007), "Reliabilism, Stability, and the Value of Knowledge," *American Philosophical Quarterly* 44 (4), pp. 343-355.
- Pennock, D., Lawrence, S., Nielsen, F. Å., and Giles, C. L., (2001), "Extracting Collective Probabilistic Forecasts from Web Games," *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 174-183.
- Plantinga, A., (1993), *Warrant and Proper Function*, New York: Oxford University Press.
- Pritchard, D., (2007), "Recent Work on Epistemic Value," *American Philosophical Quarterly* 44 (2), pp. 85-110.
- Pritchard, D., (forthcoming), "Knowledge and Understanding," in D. Pritchard, A. Millar, and A. Haddock, eds., *The Nature and Value of Knowledge: Three Investigations*, New York: Oxford University Press.
- Pronin, E., Lin, D. Y., and Ross, L., (2002), "The Bias Blind Spot: Perceptions of Bias in Self versus Others," *Personality and Social Psychology Bulletin* 28, pp. 369-381.
- Rabinowicz W., and Rønnow-Rasmussen, T., (2000), "A Distinction in Value: Intrinsic and for Its Own Sake," *Proceedings of the Aristotelian Society* 100, pp. 33-51.
- Ramsey, F. P., (1931), "General Propositions and Causality," in F. P. Ramsey, *The Foundations of Mathematics and Other Logical Essays*, London: Routledge and Kegan Paul, pp. 237-255.
- Rawls, J., (1997), "The Idea of Public Reason," in Bohman and Rehg (1997), pp. 93-141.
- Rhode, P. W., and Strumpf, K. S., (2009), "Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data," manuscript, February 2009.
- Roberts, R. C., and Wood, W. J., (2007), *Intellectual Virtues: An Essay in Regulative Epistemology*, New York: Oxford University Press.
- Rønnow-Rasmussen, T., (2002), "Hedonism, Preferentialism, and Value Bearers," *The Journal of Value Inquiry* 36, pp. 463-472.

- Sartwell, C., (1992), "Why Knowledge is Merely True Belief," *Journal of Philosophy* 89, pp. 167-180.
- Scott, D., (2006), *Plato's Meno*, Cambridge: Cambridge University Press.
- Servan-Schreiber, E., Wolfers, J., Pennock, D., and Galebach, B., (2004), "Prediction Markets: Does Money Matter?" *Electronic Markets* 14, pp. 243–51.
- Sieck, W., and Arkes, H., (2005), "The Recalcitrance of Overconfidence and its Contribution to Decision Aid Neglect," *Journal of Behavioral Decision Making* 18, pp. 29-53.
- Sober, E., (2001), "What is the Problem of Simplicity?" in A. Zellner, H. Keuzenkamp, and M. McAleer, eds., *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*, Cambridge: Cambridge University Press, pp. 13-31.
- Sosa, E., (1991), "Knowledge and Intellectual Virtue," in *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press, pp. 225-244.
- Sosa, E., (2004), "Replies," in J. Greco, ed., *Ernest Sosa and his Critics*, Malden: Blackwell, pp. 329-333.
- Sosa, E., (2007), *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*, Oxford: Clarendon Press.
- Steup, M., and Sosa, E., (2005), eds., *Contemporary Debates in Epistemology*, Malden: Blackwell Publishing.
- Stich, S., (1990), *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Cambridge: The MIT Press.
- Stiell, I. G., Greenberg, G. H., McKnight, R. D., Nair, R. C., McDowell, I, and Worthington, J. R., (1992), "A Study to Develop Clinical Decision Rules for the Use of Radiography in Acute Ankle Injuries," *Annals of Emergency Medicine* 21(4), pp. 384-390.
- Sunstein C., (2006), "Deliberating Groups versus Prediction Markets (or Hayek's Challenge to Habermas)," *Episteme* 3 (3), pp. 192-213.
- Sunstein, C., and Thaler, R., (2008), *Nudge: Improving Decisions about Health, Happiness, Wealth, and Happiness*, New Haven: Yale University Press.
- Swinburne, R., (1999), *Providence and the Problem of Evil*, New York: Oxford University Press.

- Taylor, C. C. W., (2008), "Plato's Epistemology," in G. Fine (2008), pp. 165-190.
- Taylor, S. E., (1989), *Positive Illusions: Creative Self-Deception and the Healthy Mind*, Basic Books.
- Trout, J. D., (2005), "Paternalism and Cognitive Bias," *Law and Philosophy* 24, pp. 393-434.
- Trout, J. D., (2009), *The Empathy Gap: Building Bridges to the Good Life and the Good Society*, New York: Viking.
- Whitcomb, D., (forthcoming), "Epistemic Value," in A. Cullison, ed., *The Continuum Companion to Epistemology*, Continuum.
- Whitecotton, S. M., (1996), "The Effects of Experience and Confidence on Decision Aid Reliance: A Causal Model," *Behavioral Research in Accounting* 8, pp. 194–216.
- Williams, B., (1978), *Descartes: The Project of Pure Enquiry*, Highlands: Humanities Press.
- Williamson, T., (2000), *Knowledge and Its Limits*, New York: Oxford University Press.
- Wolfers, J., and Zitzewitz, E., (2006), "Prediction Markets in Theory and Practice," Research Paper No. 1927, *Stanford Graduate School of Business Research Paper Series*.
- Wolterstorff, N., (1996), *John Locke and the Ethics of Belief*, Cambridge: Cambridge University Press.
- Wolterstorff, N., (2005), "Obligation, Entitlement, and Rationality," in M. Steup and E. Sosa (2005), pp. 326-338.
- Young, R., (2008), "John Stuart Mill, Ronald Dworkin and Paternalism," in C. L. Ten, ed., *Mill's 'On Liberty': A Critical Guide*, Cambridge: Cambridge University Press.