University of Alberta

Metacognition and Intellectual Virtue

by

Christopher Lepock ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

Edmonton, Alberta
Fall 2007

Your file Votre référence
ISBN: 978-0-494-33009-8
Our file Notre référence
ISBN: 978-0-494-33009-8

ABSTRACT

Intellectual virtues, or truth-conducive traits of intellectual character, appear to play a central role in knowledge and other forms of praiseworthy cognition. The chief problem with appealing to virtue in epistemology is the notion's lack of clarity. The relationship between a virtue and an agent's "intellectual character" has not been adequately specified. Moreover, both knowledge-generating faculties and habits for regulating inquiry can be called "virtues"; it is not clear how the two can form a single class.

My dissertation aims to resolve both problems by giving a rigourous account of the abstract structure of intellectual virtues. I argue that virtues are metacognitive capacities, or capacities for monitoring and controlling the operations of underlying cognitive processes so that they lead efficiently to the goal of significant true belief. I show that cases of non-virtues can be understood as cases where control is absent, and that this approach allows for plausible explanations of the subjective status of knowledge and the problem of metaknowledge attained by "bootstrapping". Since metacognition can be studied empirically, this notion of virtue is amenable to a detailed development. Moreover, the entire range of putative virtues can be understood as control capacities of different sorts.

In order to make these arguments, I begin by rejecting “epistemic justification”, which is generally taken as the central epistemic value despite being highly problematic, in favour of a plurality of epistemic desiderata. These desiderata are valued by how they contribute to the acquisition of significant true belief and by whether their attainment is attributable to the subject’s own efforts and powers. This understanding of epistemic appraisal allows us to give a more plausible explanation of commonsense evaluation of belief, as well as a stronger foundation on which to base the notion of intellectual virtue.

For Jay

ACKNOWLEDGEMENTS

This dissertation was written under the supervision of Adam Morton, and I have profited incalculably from his support and advice. The other members of my committee at the University of Alberta, Bruce Hunter, Rob Wilson, and Reneé Elio, were also exceptionally willing to help with my project. Jennifer Nagel guided my research into metacognition, and I have learned a great deal from conversations with her, as well as with Eric Dayton, Danny Goldstick, and Vladan Djordjevic.

Material related to this thesis was presented at the Western Canadian Philosophical Association, the Canadian Philosophical Association, and the University of Alberta. I received many helpful suggestions and criticisms from the audiences there, and in commentaries by Eric Dayton, Tom Vinci, and Patrick Rysiew.

This project was only possible because of the kind support I received from my partner, Jay Smith, and her willingness to put up with so many long days of writing. Without her family and their willingness to help with our young child, I doubt either Jay or I could have reached this point with our sanity intact.

TABLE OF CONTENTS

LIST OF FIGURES

I

THE EPISTEMIC DESIDERATA APPROACH

This dissertation is about the intellectual virtues, particularly as concerned with the acquisition of beliefs. But before we can talk about the virtues, we need at least a rough sense of how beliefs are evaluated. That will allow us to understand the place of the virtues in contributing to goods of belief, either by helping us achieve good beliefs or by transmitting their own goodness to beliefs (as in virtue ethics, on which acts are good because they are produced by virtues).

Modern epistemology has been focused on the analysis of "epistemic justification", a univocal evaluative status of overweening importance. Call this position "justificationism". The one assumption that all justificationist philosophers share—just about the *only* tenet that they all share—is that "justified" refers to an objective (though agent-relative) state, property, or evaluative status of beliefs that is of central importance. There is little reason to think that this assumption is true. Disagreements over the nature of justification run so deep that it is implausible that all justificationists are aiming at a "common target", in Alston's words (2005: 25). There is also little reason to think that epistemic evaluation features as its central standard anything resembling justification. These two considerations suggest that the focus on a single concept and its analysis is an unprofitable way of understanding epistemic evaluation. We may be better off trying to understand the interrelationships and statuses of a variety of epistemic values. The latter is of course the approach I will develop in this chapter.

1. Against "justification"

Let us first consider the controversies over justification. The many disagreements over the conditions for justification are well known and require only the briefest review. The deepest divide is between internalists and externalists. Internalists hold that whatever makes a belief justified must be within the agent's purview, in some sense of the phrase.[1] This has been the dominant position in epistemology since Descartes until the past few decades. It is generally taken to imply that it must always be possible in some sense for an agent to come to know the justificatory status of her beliefs. Principles of justification are supposed to be "regulative" (Goldman 1980: 28-30), or usable by the agent in the appraisal and revision of his own beliefs.

Externalists hold that justification may supervene on factors that are inaccessible to or outside of the agent. The most important of these are facts about the causal etiology of the belief, and in particular, whether it was formed in a manner that is objectively likely to yield truth. In general, having a justified belief does not entail being in a position to know that one's belief is justified, and agents are not guaranteed to be able to apply the principles of justification to regulate their own belief-production.

The divide between internalists and externalists often seems unbridgeable. This has led a number of thinkers to suggest that the two camps are really talking about two different evaluative properties of belief; one determined by whether the belief is well formed from the agent's perspective, and another that depends on whether the belief was formed in an objectively truth-conducive manner. The most prominent of these is Sosa, who identifies two chief evaluative properties of belief, the externalist "aptness", or production by an intellectual virtue, and the internalist "justification", a matter of perspectival coherence. "Aptness" is clearly a form of justification, since Sosa (2003a) presents it as an alternative to

[1] See, e.g., Greco 1990 for a survey of different forms of the view.

internalist conceptions of justification. (Sosa's views will be discussed in detail in §IV.1.)

Irreconcilable differences between theorists aren't necessarily grounds for divorce, but we still need some assurance that these are really different theories of the same subject matter. We could be suitably reassured if there were general agreement on what "justified" means and only controversy over the conditions for it. Not all theorists acknowledge a distinction between the meaning of the term and its conditions. Some do, however. For instance, Goldman maintains that to be justified means to have a permissible belief (1986: 59), and then argues that reliability of formation is the chief factor in determining whether a belief is permissible (1986: 103-9). Nonetheless, if there were a generally accepted, theoretically neutral characterization of the meaning of "justified", and a common practice of using the concept, we could be assured that different theorists really are arguing about a single concept. Thus although it seems sometimes that the dispute between deontologists and consequentialists in ethics is irremediable, both camps at least agree that what is at issue is what actions are permissible when, and both can build on a common, pretheoretical practice of making such evaluations. Likewise, however much epistemologists disagree over the conditions for knowledge, they can at least agree that it is non-accidentally true belief, and they can base their arguments on a common practice of attributing knowledge to agents.

Epistemologists agree no more on what justified means than they do on what the conditions for it are. Some take justification to be a deontic concept, so that a justified belief is one that violates no epistemic obligations (e.g., Plantinga, Goldman, Pollock) or is responsibly held or formed (e.g., BonJour, Chisholm). Others take a justified belief to be one that is likely to attain the goal of cognition—the acquisition of true beliefs on matters of importance (e.g., Alston in his early days, Zagzebski, Foley). Finally, others take a justified belief to be one that is supported by reasons or evidence, or that is located in the "space of reasons" (e.g., Feldman & Conee, Sellars, McDowell). (See Alston 2005: 11-5 for a detailed survey of the various positions.) And, of course, there are those who think there are several epistemic values corresponding to several of these conceptions. For Sosa, for instance, an "apt" belief is one that is appropriately truth-conducive; a "justified" belief is one that is supported by accessible reasons.

The extent of the disagreement over "justification" leads to two serious problems. First, it is not clear how exactly one is to evaluate proposed theories of justification. Do we compare the proposal with intuitions about permissible belief? or responsible belief? or what is truth-conducive? or what is a good reason for what? One would expect that the parties to this muddle would regularly be at cross-purposes. Alston proposes that this feature of the debate is the best available explanation for the extent and persistence of dispute over justification; that is, that there is no feature of belief that all the analyses of "justification" are about (2005: 21-3).

Second, many areas of philosophical analysis can ground themselves in a pretheoretical practice of using the terms in question. Thus the study of knowledge can be grounded in pretheoretical attributions of knowledge; ethics can be grounded in the pretheoretical appraisal of acts. In certain circumscribed areas, such as scientific investigation or the law, something like epistemic justification appears to play the central role that epistemologists have assigned to it. Outside of these narrow contexts, we apply a variety of different evaluative terms to beliefs. The term "justified" does occur, but when it does, it is usually clear that what is intended is the practical, not epistemic, sense of the word. For instance, Goldman cites a *New York Times* article including the sentence "[t]roops were justified in firing at what they had reason to believe was an enemy position" (2005: 134). Strictly speaking, of course, "justified" is applied to the act of firing, but the context makes it clear that "reason to believe" should be read in an exculpatory sense. It is improbable that "justified" in this sentence Goldman quotes is meant to refer to *epistemic* justification. It is more likely that what is meant is that given the exigencies of combat, especially the consequences of failing to attack an enemy position and the extreme temporal limits on decision-making, the troops were *practically* justified in their belief. To identify practical and epistemic justification is, of course, a substantive and controversial position, not an elucidation of the "common target" at which justificationists are aimed.

Thus if we are to ground justificationism in common practice, we must determine what ordinary-language expressions are approximately synonymous with "epistemically justified". In both the commonsense appraisal of beliefs and of actions, "right" and "wrong" express the most important standard. Most of the time, when we censure or try to correct a belief, we do so because it is or seems "wrong", and most of the time when we credit or adopt a belief, we do so because it is or seems "right". But unlike their counterparts for actions, these words do not mean "justified" and "unjustified"; they mean "true" and "false". (I will explore the reasons for this below.) We do often use deontic terms, as in "You have no right to assume that." (See Alston 2005: 16.) We also use terms related to evidence or reasons, terms apparently related to reliability or trustworthiness, and so forth. And then there are the many other terms—clever, intelligent, original, etc.—that do not seem to be analyzable in terms of any sort of justification. If we could know that, say, justified beliefs are permissible ones, then we could pick out the use of deontic language as the pretheoretical counterpart to the specialists' "justified", as we take the specialists' use of "knows" as a refinement of the common man's use of the same word. But if justification is *really* a matter of appropriately truth-conducive formation, then we should throw out the deontic language as misleading and take the cluster of terms about trustworthiness as expressing the core of the concept. And if justification is really about having reasons, then the talk of trustworthiness and the deontic language need to be reinterpreted in terms of evidence.

In other words, without a firmer grip on what "justification" means, it does not seem that we have reason to take any particular view of the concept as a better reflection of the underlying commonsense discourse than any other. That in turn indicates that we cannot take ordinary epistemic evaluation to support the use and centrality of the concept of "justification" at all. The available evidence favours a view on which there is a multiplicity of different epistemic values just as well as a view that "justification" is primary.

The comparative coherence of the concept of knowledge

There is a certain lack of agreement in how "knowledge" is applied among non-philosophers. This might be thought to undermine my claim in the last section that we can ground the analysis of knowledge in pretheoretical usage in a way that we cannot with justification. For instance, professional epistemologists are as far as I know unanimous in thinking that one lacks knowledge in Gettier cases. But non-epistemologists frequently allow knowledge in such cases. In a survey I conducted among students in first-year classes at the University of Alberta in 2005, 50% said that subjects did have knowledge in the standard fake-barn Gettier case. The same intuition is reported by Nichols et al. (2003), "two of three teenage youngsters" queried by Millikan (1993: 259), and some non-epistemologist philosophers I've spoken to.

This does not necessarily show that there are multiple concepts of knowledge, or that "knowledge" is ill-defined in some way. It seems more likely that (at least among Westerners)[2] some apply a shared concept more liberally than others. Nichols et al. (2003) found a correlation between restrictiveness in application of "knows" and educational attainment. (They report a correlation between restrictiveness and socio-economic status; note, however, that they use education as a proxy for SES.) It is also interesting to note that the intuitions of

[2] Nichols et al. (2003) do establish that there are cultural differences in the use of "knowledge" between Westerners and Asians. They tie this in with more general differences in how members of different cultures categorize events. For instance, East Asians attach less importance to causal etiology than Westerners do, which is reflected in a tendency to attribute knowledge in Gettier cases.

my first-year respondents did not appear to be very strong or stable; 5% answered noncommittally and 12% changed their answer before handing in their survey. These observations together suggest that as we attain a better understanding of the concept of knowledge we apply it more conservatively, or perhaps more consistently.

However, these are only hypotheses that would have to be confirmed or disconfirmed by a thorough analysis of knowledge. What is important is that there is a wide range of agreement in the pretheoretical use of "knows", and education might hone our usage. But we cannot even determine if there is pretheoretical disagreement about "justification", because the word itself is not used, and there is no agreement among the professionals about what other expressions are synonymous with it.

Ongoing and discrete acts

The sense in which beliefs are justified is frequently treated as parallel to the senses in which acts may be justified or rational. Most obviously, epistemological theories are often characterized as deontological or consequentialist. Virtue epistemology is in part an attempt to avoid the problems with traditional epistemological theories by modeling itself on virtue ethics, rather than deontology or consequentialism (as virtue theorists take other epistemologists to be—see, e.g., Zagzebski 1996, Driver 2000).

Nonetheless, there are significant dissimilarities between the appraisal of acts and of beliefs. Conventional ethics and decision theory lack any counterpart to knowledge, one of the most important standards of epistemic appraisal.[3] As

[3] Zagzebski (1996: 271-3) does discuss the possibility of ethical concepts analogous to knowledge (which would be states of subjects arising from acts of moral virtue) but she does so openly acknowledging that this is a novel endeavour. She maintains that "the structural similarity between normative epistemology and ethics breaks down at the concept of knowledge," (273) which is true but omits all the other points of breakdown.

In her (2003), Zagzebski identifies knowledge with acts of intellectual virtue. But since knowledge must be true, and as traditionally understood acts of virtue need not achieve their already mentioned, "right" usually means "justified" or "rational" when applied to acts, but it usually means "true" when applied to beliefs.[4] Ethical appraisal normally ignores whether an act successfully brought about its immediate goal (cp. Annas 2003). The same is true of the decision-theoretic appraisal of acts. With beliefs, however, success at reaching the immediate goal (i.e., truth) is a fundamental part of appraisal.

This situation seems to arise from differences in the type of entities being appraised. Ethics is primarily concerned with the appraisal of acts that are discrete events. An act can have been freely performed, and thus amenable to ethical appraisal, despite the fact that it cannot now be undone. Success or lack thereof can *now* be outside the agent's control (in any sense of the word). Thus, when we appraise an act, we are not so much interested in whether it was successful but whether it could have been expected to be successful, or whether relevantly similar acts in relevantly similar circumstances would be successful. We are concerned with whether it should have been performed, which is partially independent of whether it was successful on that occasion or not. The result is a mode of evaluation that tends to de-emphasize success on particular occasions, focusing instead on traits of the action that are conducive to overall success—the propensity for success in the long run, the states of the subject when performing the act, the skill with which the act was performed, etc.

"Believe" and "know" are stative verbs—they refer to ongoing states, not processes (Williamson 2000: 35-6). However, like "tidy" and "equitable", the states to which they refer must be actively maintained to persist over time. Thus there are certain advantages—which we'll consider below—to treating beliefs as

immediate aim (Annas 2003), this identification depends on a highly unusual understanding of acts of virtue.

[4] There are, of course, exceptions in certain contexts: one can say, "that was the wrong thing to do" when a justified act turns out badly, and one can say, "you were right to believe that" when a well-supported belief turns out to be false. And the degree to which an act is morally wrong depends in part on whether it has bad consequences; driving drunk and killing someone is much worse than driving drunk and luckily not killing anyone. Nonetheless, there is a broader divergence in meaning between the two domains from which we shouldn't let such epicycles distract us.

acts. But we must remember that *if* they are acts, they are ongoing, continuous acts, which like states persist over time. They are not just formed, but sustained. Beliefs are corrigible—or at least, when they are incorrigible, this is because of the specific situation rather than the structure of time. It is appropriate when evaluating a belief to emphasize its success at achieving its immediate goal (say, being true, or being explanatory), because an unsuccessful belief can be altered. This tends to emphasize the epistemic value of immediate success, and tends to reduce the importance of the sort of factors that are most important to the appraisal of acts.

The problem that arises for epistemic justification is that given the differences between appraisals of discrete acts on the one hand and of ongoing acts or states on the other, we cannot assume that the normative concepts applicable to each will have the same structure. And the problems with epistemic justification outlined above suggest that we do need different sorts of frameworks for analyzing the two areas.

Epistemology without justification

In the last two sections, I outlined a number of problems with the concept of epistemic justification. There is a plethora of incompatible views of justification, and we seem to have no way of deciding between them. The controversy does not just concern the conditions for justification, but runs all the way down to differences about what the term means. We have also seen that epistemic justification is closely analogous to its sister concept in ethics, but the difference between appraising discrete events and appraising states of a subject makes it doubtful that both ethics and epistemology would share the same basic evaluative concept.

I am not concerned here with the ultimate status of the concept. Rather, the point of these arguments is to suggest that chimerical analysis is not the best way to do epistemology. The alternative, proposed by Alston (1993, 2005), is that there is a variety of epistemic desiderata, and the task of epistemology is to understand these different values and their relative importance.

This is not a proposal to end debate in epistemology. Rather, the proposal is that the debate will be more fruitful if it is cast in terms of different epistemic values instead of the analysis of a single one. Steup (2004) inadvertently provides a useful example. He declares, first off, that he will address the question "Why is it that a sense experience that P is a source of justification—a reason—for believing that P?" (403) Then he observes that to answer this question, we need to make sense of what "justification" means, and he gives four possibilities. It might be

> the kind of epistemic status that:
> (1) turns a true belief into knowledge;
> (2) makes a belief objectively probable;
> (3) is as well denoted by the locutions 'having adequate evidence' and 'having good reasons';
> (4) is as well denoted by deontological locutions such as 'entitlement,' 'responsibility,' and 'permissibility,' understood in a specifically epistemic sense (404).

Steup's preference for sense (3) is not just a "terminological stipulation", because substantive controversies arise concerning the significance of internalist justification—whether it bears any weight as a desideratum in its own right and whether it is necessary for knowledge (405).

Steup's careful analysis of his own terminology will help keep him from being mired in fruitless disputation. But given this analysis, the term "justification" is ornamental. What Steup is really interested in is the question "why is it that a sense experience that P is adequate evidence—a good reason—for believing that P?" And, further, he thinks that whether one has adequate evidence or good reasons is a very important epistemic desideratum.

For another example, take the controversy between internalists and externalists. Presently, the two sides lock horns over whether the grounds for justification must be internally accessible. A growing number of epistemologists acknowledge that the two camps share different intuitions about what role justification plays, exacerbating the dilemma.

The alternative is to recognize that there are desiderata grounded in factors accessible to the subject and desiderata grounded in factors outside the subject's ken. Suppose S has reliably produced beliefs but no evidence available for them. (Perhaps S reliably remembers that *p* but cannot remember where he learned it.) Externalists will claim that S is justified but does not know he is, while internalists will claim that S is unjustified. But on the epistemic desiderata approach, we needn't take either step. We can simply observe that S's beliefs are reliably formed—which is valuable—but not supported by accessible evidence—which is also valuable. Thus we need not worry about the question of whether S's beliefs should be slotted in with the unjustified or the justified beliefs. They are less valuable than (say) ordinary perceptual beliefs, which are both reliable and supported by evidence, but more valuable than paranoid delusions, which are neither. A framework of multiple values has more resources available with which to analyze cases and systematize intuitions.

But there is still lots of room for controversy over the relative importance of these values. Alston is an epistemic-desiderata externalist and no less partisan by virtue of having become a pluralist. As we will see shortly, he maintains that reliability is a much more important feature for our cognitive lives than anything internally accessible is. An epistemic-desiderata internalist could grant that reliability is a value of sorts, but a peripheral one; having evidence is much more important to our cognitive lives. Thus the debate can rage on, but now the different schools can line up behind their preferred conception of proper epistemic activity and the roles of different statuses of beliefs in it. And by tying different views on epistemic values to different conceptions of intellectual goods, rather than unanalyzable intuitions about "justification", we may be able to better understand what is right, and what is wrong, in each school's portrayal of human cognition.

Ultimately, the merits of pluralism about epistemic values will be decided by whether it is a better framework for the projects of epistemology: understanding how we should conduct our cognitive lives, our contact with the world, the status of our beliefs, and so on. Thus this work is in part an experiment in whether intellectual virtues can be more easily understood in terms of their contribution to a family of values rather than just to justification and knowledge. In the remainder of this chapter, I will examine Alston's particular theory of epistemic desiderata, their interrelations, and their relative importance. His account is not entirely satisfactory; in ch. II we will examine its major problems and develop a version of the ED approach that fits the way that virtue epistemologists conceive of epistemic values. That will set the stage for the examination of intellectual virtue that will occupy the bulk of this work.

2. The end of cognition

Once we move from a single central epistemic value—justification—to a multiplicity of epistemic desiderata (ED), the chief task in epistemology should be to categorize the resulting desiderata, analyzing their nature, viability, interrelations, and relative importance. Alston's *Beyond "Justification"* is the first attempt at a large-scale systematization of ED. It is by no means complete, of course. As we will see, Alston eliminates some ED on grounds of inviability, and gives a detailed analysis of one group of the remainder. Beyond this, he gives a rough characterization leaving many gaps to be filled in later or by others. My presentation of the framework will be even more limited. My plan in this book is to examine the place of the intellectual virtues in our cognitive lives. Virtues themselves do not enter into the ED framework, since it is a system of epistemically valuable properties of beliefs, not of persons. But to understand what a virtue is, it will help to first understand the status of their products. That is what my discussion here is focused on.

Alston supposes that an epistemic evaluation of belief is an evaluation of how well the belief contributes to the attainment of the goals of cognition. Most generally put, the goal of cognition is "the acquisition, retention, and use of true beliefs about matters that are of interest and/or importance" (2005: 30), or "significant truth", in Kitcher's terse phrase (1992: 102). Alston writes that he cannot prove this claim, because he does not "know anything that is more obvious from which it could be derived" (2005: 30). It is virtually tautological that we should want to acquire beliefs that are useful for matters of importance to us. It is also obvious that, most of the time, to help us succeed our beliefs will have to be true. Consider wishful thinking. While there are temporary benefits from believing what one wants to be true, a habit of wishful thinking guarantees that one will be regularly disappointed on the matters one cares about most.

Falsehoods do sometimes have practical utility, but even when that is the case their value seems not to be epistemic. Believing a falsehood might bring peace of mind or improve one's chances of surviving a severe illness, but epistemic appraisal treats such beliefs as necessary evils. There is something unsettling about the prospect of believing a falsehood, even if it is a supremely useful one. It is laudable to work with approximations to the truth, rather than the truths themselves, when this simplifies the task and yields sufficiently precise results. But it is problematic to *believe* such approximations. More precisely, believing that *supposing that* $\pi = 3.14$ *would greatly simplify this computation without substantial loss of precision* is epistemically valuable; believing that $\pi = 3.14$ is not.

Moreover, there are some desires that cannot be satisfied by false beliefs. Our natural curiosity is a drive to find out the truth; it is not satisfied by explanatory falsehoods. There is something inherently valuable about true beliefs, and *epistemic* value per se requires truth.

There is, however, a certain amount of disagreement over just how truth figures in our cognitive goals. At one extreme, we have the position that we just aim at getting the truth. This view is most strenuously defended by Zagzebski (e.g., 2003). It has the problem that we are not interested in acquiring just any true beliefs. No one has the slightest interest in memorizing the Moose Jaw phone book or counting the grains of sand in a particular square foot of beach, despite all the true beliefs that could thereby be acquired. Thus it seems that some true beliefs have vanishingly little epistemic value.

At the other extreme, Sosa (2003c) entirely rejects the idea of truths having any inherent value. He proposes that our goal is actually *safety*, $Bp \rightarrow p$; we want that any beliefs we have be true ones. This does not entail wanting true beliefs—I want any illnesses I have to be mild, but I do not want to be mildly ill. But that consequence in itself is problematic, because Sosa cannot explain why we do seem to value getting some truths. Besides wanting practically useful truths, we are naturally curious about some things, and certain sorts of truths seem to have a distinctively epistemic value.

Thus *significant truth* seems to be the general structure of our cognitive end. We want significant beliefs, and we want our beliefs to be true. So we want safety, at least when the conditional is read as material:[5] for any p, we want that either p be true or we don't believe it.

These two aspects of our epistemic goal can conflict. A very discerning, demanding attitude can reduce one's number of false beliefs, but at the cost of reducing the number of significant beliefs one has; a less cautious attitude can have the opposite effects. There is thus a problem of determining how to balance these two aspects of the cognitive goal (see Riggs 2003a). I won't worry about this problem in this dissertation; we'll treat significant true belief as more valuable than merely true belief, but without determining to what extent one should aim for each.

[5] Sosa (2001) interprets it as a close-worlds conditional, meaning that $\sim Bp \vee p$ holds in all sufficiently similar worlds to our own. It is difficult to explain why this sense of safety would be of epistemic value, however, as we'll see in §II.2.

In the next section, we'll see some particularly epistemic ways in which beliefs can be significant. In the next chapter, we'll go a little deeper into what significance is in order to get a better understanding of how the ED approach can capture our practices of epistemic evaluation.

3. Alston's epistemic desiderata framework

If significant truth is the goal of cognitive activities, then (Alston proposes) ED can be categorized in terms of the contribution they make to that end. Since his aim is to provide a schema to replace the analysis of justification, the values that he discusses are an extensive collection of proposed conditions for justification. While this approach to the ED is useful for preserving continuity with the analysis of justification, it leads to certain problems with that I'll discuss in ch. II.

The first task is to cull the mass of potential ED, eliminating those that are not viable or that do not contribute to the attainment of the cognitive goal. Proposed forms of internal well-formedness not connected to truth are eliminated for the latter reason. The most notable here is Foley's (1987) proposal that a belief B is justified iff after sufficient reflection the subject would believe that B has positive epistemic status. Foley himself acknowledges that this status is not the least bit truth-conducive; thus, by Alston's standards it is not an ED (2005: 45). Deontological criteria for belief are eliminated for a combination of both reasons. Beliefs, Alston argues, are not voluntary, and thus concepts like permissibility, obligatoriness, and responsibility are not directly applicable to them. These concepts *do* apply to choices made in inquiry; so, a possible ED is the property of having been generated by permissible or responsible inquiry. But whether an inquiry is also likely to lead to truth depends on far more than just its permissibility or responsibility. Thus such a desideratum would be at best tenuously connected to truth (2005: ch. 4). Presumably, if it is an ED at all, it is of vanishingly small importance.

The remaining ED fall into four groups.

Truth

The first, most central, yet perhaps least obvious desideratum is truth—i.e., the attainment of this part of the goal itself. The centrality of truth to the ED is mirrored in commonsense epistemic appraisal. As noted above, "right" and "correct" usually mean "true", "wrong" and the like usually mean "false", and we endeavour to correct beliefs primarily and most thoroughly when we think them false.

Other goals of cognition

This category contains certain ways for beliefs to be significant that are not too closely tied to specific interests but that are of particular cognitive importance. They include understanding, systematicity and coherence, and explanatoriness. We should add wisdom, and novelty, which contributes to the expansion of human knowledge on the whole (see Zagzebski 1996: 182-3).

These goods do not really belong to individual beliefs. Systematicity and coherence are properties of sets of beliefs; explanation is a relation between beliefs; understanding and wisdom are probably best understood as properties of agents, since they might involve the agent's capacities, virtues, and so forth. It seems that individual beliefs can be more valuable by virtue of partaking in one of these goods (for instance, by being part of a coherent set), or by virtue of contributing to its achievement (by being one of the beliefs that ties the whole set together, or by allowing one to understand some aspect of reality).

Alston argues that these goods are epistemic values because they are really only valuable when the beliefs that have them are true (or at least mostly true). There is obviously a rich structure of interrelations in this area, the examination of which would distract from my main purposes here. Some of these values, particularly understanding and wisdom, will come up from time to time, but we will generally have to rest on an intuitive sense of what they are.

Characterizing the cognitive goal as significant truth and saying that these ED are forms of significance does not give us a very unified account of them, but we might be able to do better. We might take the cognitive goal to be something like having an accurate model of reality that includes its significant aspects and permits successful action. Then it might be possible to understand coherence, explanatoriness, etc. as valuable properties of such a model. However, I will not attempt to establish this (nor even try to state it rigorously). I mention it only to show that we might be able to get a better explanation of this group of desiderata than Alston provides.

Directly truth-conducive desiderata

Weaker, and thus less important, than truth are what Alston calls the "directly truth-conducive" desiderata. These entail the probable truth of the belief (rather than presupposing truth or probable truth, as with the group just discussed). They include having adequate evidence for one's beliefs (where "adequate" is understood as meaning, "rendering probably true"), and formation by a reliable, properly functioning, or virtuous process.

Here Alston does give a detailed examination of the relevant interrelations inside the group. He argues that they are all variations on the same central value, production by a reliable process—a process that generates beliefs that are mostly true. He argues that evidence E is adequate to support belief in p iff the process of believing that p on grounds E is reliable. Properly functioning and virtuous faculties are truth-conducive only insofar as they are reliable; Alston maintains that the additional features they have beyond reliability do not contribute to the goal of attaining true beliefs, and thus do not add epistemic value to the beliefs they produce. (In ch. II, we'll see why this assessment of virtues is problematic.)

Reliability is of strictly lesser value than truth, since it is valuable only inasmuch as it has a high probability of yielding truth. The likelihood that a belief is significant is not relevant to the desiderata in this group, but an interesting question is whether there is a special value attached to a belief by virtue of its being formed in a way likely to yield significant beliefs. "Coherence-seeking reason", for instance—Sosa's term for the group of processes that draw out implications of beliefs and find contradictions and explanatory relations among them—would be reliable when conjoined with reliable input faculties, but it seems that it might be even more valuable than the input faculties because its products are likely to fit together into a coherent whole. More generally, it seems that to say that a person has an intellectual virtue is to say not just that they can attain truths, but that they can attain important ones. An intelligent person is not just someone who can solve lots of problems, but someone who can solve certain sorts of problems that are of special importance.[6] A creative person is someone who finds novel solutions to problems of interest and importance, not just someone who comes up with unheard-of true beliefs.

In this chapter, I'll ignore this possibility. We will come back to it in ch. IX when we consider how intellectual virtues are valuable. It is plausible, after all, that at least some virtues might be especially valuable because they contribute in special ways to our cognitive goals or to human flourishing.

Indirectly truth-conducive desiderata

One important group of putative accounts of justification are those involving higher-order access to a belief's grounds or status. These include access to the evidence for the belief; and access, well-grounded belief, or

[6] The sorts of problems on IQ tests do seem to have some special importance for life in modern, Western society, since IQ is strongly correlated with academic and occupational success therein (see Ceci 1996a, Hunter & Schmidt 1996). It's an entirely different matter whether they have any further significance (and thus whether IQ measures anything deeper than a capacity to succeed in our society).

knowledge of the epistemic status of the belief.[7] There is a rich structure of different types of access and different epistemic statuses of meta-beliefs, which it is prudent not to examine too closely here. To keep at least a little rigour in the discussion, however, let us suppose that the access in question involves either (a) conscious awareness of the grounds or epistemic status of the belief, or (b) a doxastic state—a belief or belief-like state, a state to which epistemic evaluation can legitimately be applied—whose content is the grounds or epistemic status of the belief. Call the belief "consciously or doxastically accessible", or CDA.

A belief whose truth or reliability is accessible to the agent makes a more secure contribution to the truth-goal than one whose status is inaccessible. It is not as easily abandoned in the light of apparent counterevidence when its status is known; it is easier to apportion resources in inquiry when one has access to what is already true or probably true; etc., etc. Having adequate evidence for one's beliefs is more valuable than just having reliably formed beliefs because evidence, being evident, is *prima facie* consciously accessible. So evidentially supported beliefs have two valuable properties—reliability and CDA.

CDA makes much less of a contribution to cognitive goals than success or likely success. Certainly, CDA does not contribute to significant truth when conjoined with beliefs that are not true or reliably formed. A belief that is likely to be false reflects even more poorly on the agent when she knows it is probably false. More importantly, on this conception of the goals of cognition there is no value in having access to grounds that are not at least reliable indicators of truth. Merely having evidence that one thinks is adequate, but which in fact is not, is not epistemically good on Alston's account.

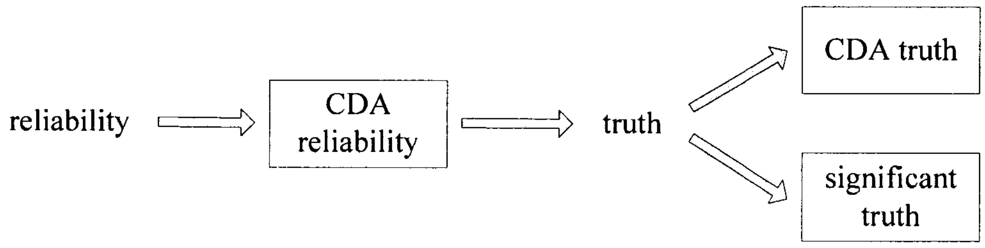
The position of a CDA belief is akin to that of a belief that can be usefully generalized upon, or a general claim that can be applied to form beliefs about

[7] A third Alston considers is the capacity to construct a defense of the probable truth of the belief. This does not seem to be best understood as an epistemic desideratum, since it depends not just on the status of a person's beliefs but also how articulate he is. Its purely epistemic value probably reduces to access to evidence and epistemic status, since this access is what permits a sufficiently articulate subject to mount a defense of a belief.

specific instances (for an agent with the appropriate inferential capacities). Like CDA, those properties of belief tend to lead to the acquisition of more truths and fewer falsehoods in the belief-set; the only difference is that they have this status by virtue of the relation between their content and the agent's belief-forming capacities. Another case of this worth mentioning is the property of contributing to understanding. Understanding a subject seems to imply having a capacity to form reasonably trustworthy beliefs about it (see Riggs 2003b: 219-20). Understanding how a car works implies being able to determine the effects of changes to the car, say, by specially tuning the fuel injection or the muffler. It also involves being able to determine the causes of various sorts of malfunctions.[8] Understanding a language implies being able to interpret novel utterances expressed in it. Thus beliefs that contribute to understanding are indirectly truth-conducive as well as intrinsically valuable.

The importance of any particular property of belief will vary depending on the precise situation at hand (see 2005: 170-84 for discussion), but we can still describe (as I have above) the general relationships between different structures. The resulting partial ordering of values is summarized in figure 1. The arrows point from strictly less important values to strictly more important ones. Arrows

Figure 1. Alston's ED framework



[8] Of course, it does not entail being able to do this a priori. Someone who understands how a car works knows how to examine and test the car in a way that would determine what the problem is.

are missing where we have no reason to specify relative importance. The boxes signify areas with internal structure that I have not elucidated here, but that is entirely contained within the appropriate place in the diagram. Thus any form of CDA added to reliability is more valuable than reliability alone, although there may be much to say about different types of CDA. Relative length of arrows does not indicate relative differences in value. Since the diagram illustrates Alston's views, it ignores indirectly truth-conducive desiderata he does not discuss as well as direct truth-and-significance-conduciveness.

II

AN AXIOLOGY FOR VIRTUE EPISTEMOLOGY

In the previous chapter, we examined Alston's framework for categorizing epistemic desiderata. Since Alston is concerned with showing how we can do epistemology without epistemic justification, he does not discuss knowledge. Knowledge is surely an epistemic desideratum, and it should be included in our framework. Two problems arise when we try to fit it in. First, evaluations of knowledge do not involve significance. So it may not be clear how knowledge fits into an axiology that takes the cognitive goal to be significant truth, not just truth. We will see that we can explain this without changing Alston's framework. That is not the case with the second problem, which is to explain why knowledge is more valuable than true belief. To solve this problem, we will have to revise our categorization of epistemic desiderata. The resulting framework will capture epistemic values as virtue theorists conceive of them.

1. Significance, truth, and knowledge

As we saw in ch. I, Alston's formulation of the end of cognition as significant truth finds an appropriate middle ground between the view that the goal is the acquisition of truths full stop and the view that the goal is safety, $Bp \supset p$. We want our beliefs to be safe—true beliefs are better than false ones—and there are many matters on which we want beliefs (which should then be safe).

The trouble with making significance part of the cognitive goal is that much epistemic evaluation does not involve significance. If I memorize a randomly chosen number from the phone book, I know it despite its banality. On traditional views of epistemic justification, a belief can be justified without serving any practical purpose. Grimm (forthcoming) maintains that it is hard to see how we can base our analysis of forms of appraisal that do not involve significance on their being conducive to a goal that does.

So the question we should address is why there should be forms of epistemic appraisal that do not involve significance. To answer this, we will have to look in more detail at what goes into these two sides of the epistemic goal.

Truth

Let me begin by indicating what assumptions I am inclined to make about truth. None of these is particularly controversial, at least not as far as the analysis of virtue goes. Thinkers who reject them have their own problems to deal with. (Some thinkers prefer not to have to address epistemological problems, and will go to great lengths to avoid it.) I won't defend my weak assumptions about truth here, because they are widely discussed elsewhere,[1] and because I have every expectation that readers who do not share at least weak versions of them will not have made it through the first chapter.

For the characterization of our cognitive goal as in part "truth" to be meaningful, we must assume a minimal correspondence theory of truth (Alston 2005: 31). To say that we want to have true beliefs is to say that whether the goals and interests of cognition are attained depends not just on the cognitive processes, or the agent's community, but on the common world in which we all live, and whether the world is the way our beliefs represent it as being. What is particularly important is that the way the world is be sufficiently independent of beliefs and social practices for truth to play a serious role in explaining cognition. One might accept that we want our beliefs to correspond with the facts, but then claim that since the facts are socially constructed by communities, the appeal to truth plays only a shallow explanatory role.

[1] See, e.g., Alston (1996), Russell (1912: ch. 12), Sosa (2003b), and Brown (2001).

We also must assume that the way the world is, is sufficiently independent of our practical interests for it to be intelligible to evaluate the truth of a belief independently from its practical importance to us. This assumption is not necessary for addressing the question at hand—if "truth" were just a special form of significance, we wouldn't have to explain why knowledge-appraisals don't involve significance—but it's worth mentioning in the interest of full disclosure.

Finally, and crucially, I assume that since beliefs are about a common world, their truth-values are the same for all believers. I take it this follows from a sort of disquotationality principle for belief; i.e., for all p, Bp is true iff p. This assumption does not imply that there are no individual or cultural differences in how things seem, or what is labeled "true", or what can be conceptualized, expressed, or comprehended. It is important because it introduces a common element into the appraisal of different believers. If S's belief that p is true, then so is T's belief that p. Likewise, if process P is truth-conducive in environment E with respect to propositions in F, then this is true whether P is instantiated in subject S or subject T. There is no idiosyncrasy to what beliefs are true.

When we look at the different ways that beliefs can be significant, what we see is that whether a belief is significant depends highly on one's goals, interests, and other beliefs. In other words, while truth is not meaningfully relative to believers or groups thereof, significance is. We can see this by looking over the chief ways in which beliefs can be significant.

Practical significance

Now let us look at the different manifestations of significance. Most obviously, beliefs can have practical value—they can help us satisfy our various goals and interests. But what beliefs have practical value depends on what our goals and interests are. The beliefs that I need to reach my goals are quite different from the beliefs that a child labourer in an Indonesian factory needs.

One can expect that there will be beliefs in this category that are significant for everyone. At a sufficiently abstract level, one can find invariants in all human lives. Most obviously, we all have to live with other persons, and there are certain similarities in the ways that we all must deal with others in order to get what we want, broadly construed. There are similarities in what constitutes flourishing or success for all persons. Thus there are certainly beliefs that are significant to all human beings. Wisdom presumably consists in part of having such invariantly significant beliefs.

Of course, this is only a class of beliefs that would be significant for every human being. Other cognitive agents (actual or possible) might lead sufficiently different lives that they would need substantially different beliefs. Even for humans, it isn't plausible that these basic ethical beliefs are very specific or very numerous. While there are surely certain things that everyone in a relationship must know, there are many things that I need to know in order to thrive in my relationship that others do not. There is as much diversity in human lives as there is commonality, and this implies a wide range of beliefs with practical significance for some and not others.

Epistemic significance

We saw in the previous chapter that there are certain epistemic goods that we will regard as ways for beliefs to be epistemically significant—being part of a coherent set, explanatoriness, contributing to understanding, and the like. Recall that for some of these a belief can be significant by virtue of possessing the property or by virtue of lending that property to other beliefs; for instance, a belief can be part of a coherent set or (even better) can be what makes a belief-set coherent.

Here, too, there are certain invariants, but not many. Obviously, with the exception of self-contradictory beliefs, whether a belief is part of a coherent set depends on the rest of the set. What is coherent for S to believe may be incoherent for T. What contributes to understanding can also vary between agents. To understand (say) why p is a logical truth, S might need to see a proof. But the symbolic proof may not do the trick for T; T might need to draw an abstract diagram. I will not comment on the complex question of whether explanatory relations are invariant or relative to belief-sets; which way that turns out will not substantially affect the conclusions I will draw below.

One important way in which a belief can be truth-conducive is by leading to other true beliefs. Good scientific theories allow one to make predictions, explain specific phenomena, and lead to promising avenues for new research. This might seem invariant, but only because we implicitly understand it as relativized to the scientific community. While general relativity is a fruitful theory, it cannot lead to any new true beliefs for those who lack the mathematical acumen to draw out its implications. As I understand it, different string theories make different predictions (and thus at least some of them might lead to new true beliefs) but not at energies that we can actually observe. Until we build better particle accelerators, these theories can only have explanatory value for us. So despite certain exceptions there is a wide range of variation among what beliefs are epistemically significant.

Natural curiosity

In addition to having practical interest in some questions and being able to attain epistemic goods from others, we are also naturally curious about some things. From basic science to reading gossip magazines, much of our inquiry has no purpose except the simple desire to know. Obviously, we're selective in what we're curious about. But it seems that a belief can be significant by virtue of exciting our curiosity.

Grimm (forthcoming) makes a great deal of progress towards an understanding of distinctly epistemic curiosity. He observes that there is very little that is invariant about the subject matters that excite curiosity and puzzlement. Rather, curiosity can be characterized by type of question; there are certain question types for which we naturally want to have answers. He discusses two:

(1) What is he/she doing?

(2) Why are things this way rather than that?

These can readily be seen to excite our curiosity. When we see someone doing something, we generally want to know what it is. Likewise, when things are a certain way, we often for no apparent reason want to know why. The two overlap to a certain extent, since an adequate answer to (1) should not just indicate what the agent is doing, but provide some account of why that agent is doing that. The situation is quite different with, e.g., "How is this done?" It seems that when we want an answer to this, it is for practical reasons. Otherwise, we are quite happy to let people do things without our knowing how. If one hears an old man yelling in an empty room, one will almost certainly want to know what he is doing and why he is doing it. But only those with some particular interest in performance will want to know how best to prepare to play Lear.

We could easily add a third question type that excites curiosity:

(3) What kind of thing is this?

As with the others, we always want to have some sort of answer to this. One will almost always want to find out enough about a strange object to at least assign a rough type to it. With these three questions, we have grammatically correct versions of the "Wha doing?" "Why?" and "Whassat?" with which young children pepper adults.

Nonetheless, although this allows us to give an abstract characterization (or a start on one) of what we are curious about, there is still tremendous variation in what *beliefs* will be significant by virtue of being answers to such questions.

On (1), we are obviously more interested in what certain people are doing than others. Anglo-American philosophers might track faculty moves among other Anglo-American philosophers, but few keep track of faculty moves among anthropologists or Central Asian philosophers. (Grimm concludes from this that our interest in questions of this type is only partially epistemic.)

On (3), what counts as a satisfying answer depends on how finely one categorizes that sort of thing, how rich one's beliefs about it are. One person might be satisfied to know that this music is piano music; another would want to know that it is Bach's Goldberg Variations, performed by Glenn Gould, and still might be curious as to whether it was the 1955 or 1981 recording.

As I noted above, I won't delve into the question of whether explanatory relations are invariant. Certainly, the effort we are willing to put into answering a question of type (2) depends on how fruitful it is. I don't care to know why my laptop is where it is rather than ⅛" to the northwest. If I could trace through the exact physical sequence involved when I put it on the desk, I could find an explanation for that fact. But since this wouldn't teach me anything else I find significant, it's not worth doing—whereas, like all academics, I am willing to put in a great deal of work to solve problems that I do think are fruitful. But the fruitfulness of a belief depends, as we saw, on one's capacities to draw out other beliefs from it; and that is highly variable.

What these observations show is that even if we can characterize the types of questions about which we are curious, this does not identify a class of beliefs that are significant for everybody.

The moral for interpersonal appraisal

This consideration of the types of beliefs that are significant show that what is significant is highly variable between agents. Significance is relative to belief-set, interests, environment, belief-forming capacities, and the like. On the minimally realist approach to truth we have assumed, truth is not relative to belief-set, interests, and so forth. For any set of diverse agents inhabiting a common world, a belief in p will have the same truth-value. But the degree to which p is significant for different agents in the set can vary immensely.

Thus it makes sense that community-wide standards of epistemic appraisal, like knowledge, should favour truth-conduciveness over significance. A trait can be truth-conducive or a belief true full stop, but can only be significance-conducive or significant for certain agents and not others. Thus the information carried by appraisals purely of truth and truth-conduciveness is more widely applicable than that carried by appraisals involving significance; and appraisals of the former sort are (more or less) absolute, while appraisals of significance are always relative to interests, capacities, and the like.

So suppose that our central form of appraisal was "sknowledge", which refers to significant knowledge. "Sknowledge" would almost always have to relativized to agents or classes of agents; we would have to say "S sknows-for-x that p", as in "S sknows-for-Hitchcock-fans why Hitchcock makes a cameo in each of his movies," or "S sknows-for-campaign-strategists that the fundamental issue in this election is the state of the economy". It should be obvious that there are many reasons why this would not be a happy state of affairs.

By saying that S knows why Hitchcock makes a cameo in each of his movies, we convey (among other information) that *if* you care to know and you can get S to tell you, S is a good source. Likewise, if we say that S is a reliable source for "Star Wars" trivia we put the hearer in a position to make use of S's knowledge if he so chooses; we leave it up to the hearer to decide whether S's reliably true beliefs matter for him.

The unpredictability of significance

Even once we relativize our appraisals to a particular individual, it can be impossible to determine what beliefs are significant. This is partly because you often cannot know what will be significant to you at future times, given your future beliefs, interests, and so forth. Significance varies over time for a single agent just as it varies between agents.

Moreover, even given a particular agent's interests, it can be virtually impossible to predict what beliefs will satisfy those interests. One typically cannot know ahead of time what new beliefs will solve a particular problem. Thus solving a problem often involves collecting large amounts of information most of which will turn out to be unimportant. The discovery of penicillin, for instance, arose from the chance observation of a contaminated Petri dish in a notoriously messy lab. Sherlock Holmes must carefully observe every tiny detail, because he has no way of determining what apparently harmless point will turn out to be an important clue. The RCMP assiduously collects information about the sale of large quantities of fertilizer. Only a vanishingly small proportion of such beliefs have any importance whatsoever. But since the Mounties can't tell ahead of time which purchases of fertilizer might indicate a bomb plot, they are stuck keeping track of all of them.

So here is another case where it is useful to evaluate truth and truth-conduciveness separately from significance. We may not be able to tell whether a detail of the crime scene or a fertilizer purchase is important, but we can take measures to ensure that our beliefs about them are accurate; and if they do turn out to be important, it is vital that we represent them correctly. Thus in inquiry we want to amass knowledge, most of which we need no longer retain after we've found our solution.

The unpredictability of significance seems to be the reason why our faculties are not particularly efficient at filtering out insignificant beliefs. One's memory is full of bits of information and fragments of episodes from long ago; one's perceptual experience full of unimportant details. Even our practices of inquiry are not very focused on significance. We all lazily read articles in the newspaper that we know are on subjects in which we have little or no interest. My father used to count the ceiling tiles in his church to ease the tedium of the

Latin mass. There's not much cost to such practices, and the potential for real gain; habits and faculties that are too focused on acquiring ostensibly significant beliefs are very likely to lead one to miss unexpectedly important truths.

None of these conclusions should be taken to indicate that appraisals regarding significance are not important. (We will see several examples of important appraisals involving significance over the course of this thesis.) Rather, it is to say that appraisals not involving significance are more generally applicable, and thus have a special function in epistemic discourse even though they only carry information about one axis of our cognitive goals.

2. The value problem

We have now seen Grimm's objection to Alston's project of deriving epistemic norms from epistemic desiderata. A second group of objections argues that purely instrumentalist theories like Alston's cannot capture epistemic intuitions.

It seems intuitively obvious that as a rule, knowledge is more valuable than mere true belief. That is, it is better for a belief to be known than to be merely true; knowledge is a greater desideratum than truth. However, there is no room in Alston's framework for knowledge as an ED superior to merely true belief. Recall that we can locate the combinations of truth and access to grounds, and truth and significance, as distinct values. But it's hard to see any reason why reliably produced truth would be any more valuable than unreliably produced truth. Reliability only has value by virtue of being likely to be true. Significance is a value of its own, and access to grounds at least guards against mistaken rejection, but a true belief is no better because it is also likely to be true. Thus reliably formed true belief seems to be no better than unreliably formed true belief.

This problem of explaining how knowledge can be better than accidentally true belief is important enough to sometimes be called just “the value problem”. It is generally seen as a problem for “epistemic value monism”, the view that truth is the only fundamental epistemic value, and all other desiderata are valuable only by virtue of leading to it. We cannot accuse Alston of value monism, since he endorses epistemic goods like coherence, understanding, explanatoriness, and the like. But this particular plurality of values won’t help the problem. Knowledge differs from merely true belief in part by virtue of its etiology—it is non-accidentally true, safe, reliably formed, or what have you. The factors that distinguish knowledge from true belief do not seem to be cognitive goals in their own right. (The one possible exception is access to the grounds of the belief, which we will discuss shortly.)

We might be tempted to think that safety, $Bp \rightarrow p$, is the goal that distinguishes knowledge from mere true belief. If we read the conditional as material implication, it admits accidentally true belief as safe. We could interpret safety as involving a close-worlds conditional, so that it means that $\sim Bp \vee p$ holds in all sufficiently similar worlds to our own (Sosa 2001). But then what we have to explain is why a true belief should be more valuable because it is also true in similar situations. It is not clear why a belief would be less valuable because it might have been false, but isn’t.

Furthermore, knowledge has nothing to do with significance; unimportant, unexplanatory, and unexplained truths can be known. So it’s implausible that what makes beliefs known would gain its value from any goal other than truth. Alston does not seem to be in any better position with respect to the value problem than value monists.

Non-instrumental desiderata

One seemingly natural route to a solution is to try to explain why it is better for true beliefs to be reliably rather than unreliably formed. This won’t do the trick because there are well-known examples of apparently reliably formed true beliefs that are not knowledge. The most famous is probably BonJour's (1980) case of Norman the unwitting clairvoyant. Norman has reliable clairvoyant powers, but no reason to either believe or disbelieve that he has those powers. Suppose Norman believes, on the basis of his clairvoyance, that the President is in New York City, and this is true. BonJour argues—and very many people agree—that Norman does not know that the President is in New York, even though his belief is reliably formed.[2]

Plantinga (1993a: 199) provides another illuminating example. Suppose that one has a brain tumour that causes only one belief, the belief that one is going to die; furthermore, the tumour is fatal. The belief that one is going to die is reliably formed,[3] but nonetheless not knowledge.

Here is a third, due to Hilary Putnam. Suppose that the Dalai Lama is infallible. Then believing everything the Dalai Lama says is perfectly reliable. But suppose that the believer's only reason for trusting everything the Dalai Lama says is that the Dalai Lama says he should (1983).

And a fourth of my own. The Shining Path to True Science, a group of highly educated radicals, has seized control of a small country and dedicated themselves to wiping out the superstitions and other false beliefs about the empirical world that are accepted by the general population there. The benighted are sent to re-education camps where they are abused in various horrific ways that the SPTS insist do not constitute torture, but that have the effect of breaking the

[2] Note that whether Norman has knowledge actually depends on details of the case that are not given here. As a detailed analysis of the case in chapter 4 will show, it is possible for Norman to have knowledge on the basis of his clairvoyance. There is a common tendency (perhaps due to "a bias against clairvoyance", as Bernecker forthcoming, maintains) to read the case in the worst possible light and thus agree with BonJour's conclusion on the matter.

[3] The tumour's reliability does not depend on the fact that the belief and the truth of its content have the same causal origin, although that is the version of reliabilism the counterexample was originally designed for. As long as the possible situations in which the tumour causes the belief are a subset of those in which it is fatal and the tumour causes no other beliefs, all the beliefs it produces will be true.

victim's spirit and leading them to believe anything their captors tell them.[4] Once this has been accomplished, the prisoners are taught basic scientific facts, all of which are indisputably true. Forced re-education is a reliable process for acquiring true beliefs, but intuitively, not one that yields knowledge.

The moral of these thought-experiments is usually stated in terms of justification and similar concepts. BonJour maintains that reliable processes do not generate justified beliefs (1980); Plantinga (1993b) uses the example to argue that only properly functioning faculties generate beliefs with "warrant", a necessary condition for knowledge. It is a short step to putting the moral in terms of epistemic desiderata; the examples show that some reliably formed beliefs are more valuable than others. As with knowledge, it is hard to see how to account for this difference in terms of indirect truth-conduciveness. Reliably formed perceptual and mnemonic beliefs are better than beliefs produced by reliable tumours, whether or not we have conscious access to their grounds or meta-beliefs about their status. One reason why properly functioning faculties and virtues are invoked in some theories is precisely to account for the variable value of the products of reliable processes. But it doesn't seem as if a belief is less likely to satisfy our goals because it was produced by a reliable tumour than by (say) reliable perception. At the very least, we need a more refined understanding of epistemic desiderata to answer the question.

The internalists return

We might also be able to solve the problem by moving back towards a classical internalism. Internalists traditionally argue that rational cognition *requires* access to epistemic status.[5] In §I.1, I argued that disputations over the

[4] This appears to be an effect of the resource-depletion caused by forms of torture like sleep deprivation. It seems that victims of the Soviet secret police and the Cultural Revolution sometimes believed the fabrications to which they were forced to confess (Gilbert 1991: 111).

[5] Internalists of a certain stripe, at least. A second important type of internalism is more or less defined by Davidson's (1986) dictum "only a belief can justify another belief". This sort of

nature of justification are fruitless. But I also noted that the epistemic desiderata approach does not dissolve debates so much as reconfigure them. Rather than saying that justification requires internal access, we could hold that a property must be accessible to reflection in order to be an epistemic desideratum.

Here is a rationale for such a move. Epistemic desiderata are valuable inasmuch as they are *regulative*, as they provide first-person reasons for preferring one belief to another:[6]

> Why, the internalist will ask, should a reason that is outside the cognitive grasp of a particular believer nonetheless be taken to confer [an ED] on his belief? Is this not indeed contrary to the whole idea of [an ED], which surely has something to do with selecting one's beliefs responsibly and critically and above all *rationally* in relation to the cognitive goal of truth? How can the fact that a belief is reliably produced (or indeed any sort of fact that makes a belief likely to be true) make my acceptance of that belief rational and responsible when that fact itself is entirely unavailable to me? (BonJour 2003: 27, e.i.o.)

(I've replaced each instance of "epistemic justification" with "an ED", to make BonJour's argument appropriately general.) "Available" does not just mean "discoverable" or "learnable", but "available to introspection":

> [i]f the question is whether I have good reasons for my beliefs, then the answer must appeal to reasons that I genuinely have and which are thereby available or accessible to my reflection (2003: 175).

For BonJour, agents have a duty to reflect on their belief-formation, and they

position leads directly to one or another brand of coherentism. Coherentism does not entail access to epistemic status, since one may be unable to determine whether one's beliefs are coherent (Dancy 1985: ch. 9, BonJour 2003: ch. 3). Internalists of this sort reject the idea that epistemic status can supervene on non-epistemic properties of belief like the frequency with which the process that generated them are correct (see, e.g., Putnam 1983). Tackling positions of this sort in this thesis would take us too far afield, so I will set them aside.

[6] See also Goldman (1994) for a general characterization of internalism as the demand that epistemic principles be regulative.

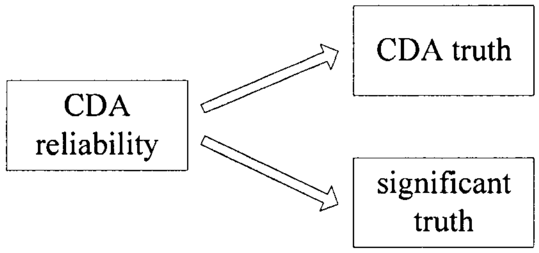
violate that duty by holding beliefs for which reflection cannot find adequate reasons. We can avoid the talk of duties but get essentially the same result by supposing that it as *constitutive* of ED that they be accessible to introspection. Qualities like mere reliability that are not necessarily accessible to reflection or introspection might be good, but not an epistemic one.

BonJour argues for this view of epistemic goodness by maintaining that it is necessary for responding to skepticism. Thus a discussion of skepticism, and the relative merits and demerits of internalist and externalist approaches to it, might be thought necessary here. It would, however, be a long digression, mostly peripheral to the rest of this thesis, and would recapitulate superior work already existing in the literature.[7] Let us instead look at two more prosaic questions: can an internalist system of ED explain the value of knowledge, and can it make sense of our epistemic evaluations?

Internalist knowledge

The resulting framework is illustrated by figure 2, which results from figure 1 after deleting all desiderata that may not be accessible to introspection.

Figure 2. An internalist ED framework



Since the resulting scheme makes use of two axes, it is no longer purely instrumental. Recall that "CDA" stands for "conscious or doxastic access to

[7] See, e.g., BonJour & Sosa (2003) and Greco (2000a).

grounds". Let us assume that access to the grounds of a belief need not be infallible. In CDA truth, for instance, the subject has introspectible reasons for thinking the belief true, but they need not be conclusive. A more complete account might allow for a distinction between, say, reliable access and conclusive reasons, but that would only complicate matters.

This scheme does allow for an explanation of why some reliable processes are less valuable than others. The reliability of brain tumours, clairvoyants, and the like is (presumably) not CDA, and thus not epistemically valuable at all. We also have an explanation of the value of knowledge. Knowledge is at least CDA truth; it is more valuable than mere true belief because in knowledge, the truth of the belief is accessible to reflection.

Gettier problems are a bit of a problem for this account. In a Gettier case, the subject's belief is supported by appropriate (accessible) evidence and true. But since it is only true by a remarkable coincidence, it fails to be knowledge. Whether a belief is gettiered is not accessible to reflection. Thus, CDA truth is not necessarily knowledge.

This might just be taken to be an interesting consequence. Perhaps knowledge is more valuable than mere true belief because in knowledge truth is accessible to reflection. Gettiered CDA truth is not knowledge, but also not any less valuable. While some authors do maintain that knowledge is *always* more valuable than non-knowledge, this is controvertible. *Prima facie*, it does seem rather odd that a belief would be more valuable by virtue of being ungettiered. So perhaps a case could be made for this conception of the value of knowledge. Nonetheless, I won't try to work out the details, because there are clearly untenable consequences of adopting this framework.

Internalist desiderata

While an internalist framework may have some intuitive appeal, it does not reflect our practices of epistemic evaluation. We often distinguish between beliefs on the basis of factors that are not accessible to the believers. Suppose Stanley acquired the belief that the Battle of Hastings occurred in 1066 many years ago on the basis of solid evidence. Since then, he has learned nothing that would suggest that he acquired the belief from bad evidence or that the belief was false, but he has also completely forgotten where he acquired the belief and what evidence he once had for it. Stanley no longer has conscious access to the grounds for his belief, but we are nonetheless inclined to say that he knows that the Battle of Hastings was in 1066 (Goldman 1994: 309-10).

Memory has a rich phenomenology that we make use of in deciding whether to accept an ostensible memory as veridical. Memory retrieval comes with an experience of greater or lesser fluency—the speed with which the information can be retrieved, the persistence of the retrieved information, and the amount of associated information that comes with it (Benjamin & Bjork 1996). Since we use fluency as an indicator of whether an ostensible memory is to be trusted, we might suppose that an experience of sufficient fluency might constitute a consciously accessible reason for trusting a memory even after its source has been forgotten. But consider Sally, who when confronted with the exam question "When was the Battle of Hastings?", noticed that it was 1:06 PM, wrote down '106', and then added another '6' on the grounds that years with three digits look funny. Sally now seems to remember that the Battle of Hastings occurred in 1066 with exactly the same feeling of certainty as Stanley. But, of course, she does not know it.[8]

Here, the internalist cannot plausibly say that the two beliefs are equally valuable. Sally's is clearly inferior to Stanley's. This is the case even when neither has knowledge. Suppose both believe that the Battle of Hastings occurred in 1065; Stanley, because of a typo in an otherwise trustworthy book; Sally, because her watch said it was 1:06 and 52 seconds, and 10652 is a *very* funny-looking year. Neither has knowledge, but Stanley's belief is still clearly superior.

[8] Cp. Greco (1990)'s arguments that whether one believes responsibly is not internally accessible.

Thus an internalist ED framework of the sort we have considered ultimately cannot explain our practices of epistemic evaluation. The general moral here is that the status of a belief depends on its causal history, and causal history is not always accessible to reflection. We will have to take another approach to explain the value of knowledge.

3. The virtue-theoretic response

In this section, I will lay out the virtue-theoretic response to the problem. Loosely, this is that the value of a desideratum is increased when it is attained by one's own efforts and powers. That allows us to rank desiderata on two axes: truth-conduciveness and the degree to which the truth-conduciveness is attributable to the subject.

First, we need to lay some groundwork. The objection that reliabilism cannot solve the value problem acknowledges that reliable *processes* are more valuable than unreliable ones. A reliable espresso maker is more valuable than an unreliable one. However, Zagzebski (2000) argued, the value of the reliable process is not transmitted to the product. When an unreliable espresso maker does produce a good shot, the result is no worse than if it were produced by a reliable machine, only rarer. This led Zagzebski (2000) to argue that the difference in value arises because knowledge is motivated by the desire to attain the truth, and merely true beliefs are not. The value of an act's motivation, she argues, is transmitted to the value of an act.

This response won't work because well-motivated true belief is not knowledge and as a rule not as valuable as knowledge. Suppose Hubert is deeply motivated to believe the truth as to whether it will rain tomorrow, and that is why he eschews the weather report and consults chicken entrails. The resulting belief (supposing it turns out true) is not much more valuable than a guess, and certainly less valuable than the result of using a reliable source would have been.

The more sensible way to approach the problem is to allow that the etiology of a belief can affect its value. This is probably what would have to be done in any case, since knowledge is distinguished chiefly not by the final belief-state, but by its history. In general, it is possible for an object's history to influence its value. A dress that belonged to Princess Diana is more valuable than an exact duplicate just because of having been owned by her; a painting by Rembrandt is more valuable than an exact duplicate by virtue of having been painted by him.

As I noted in §I.1, it is common among virtue epistemologists to treat belief as an act. Zagzebski (2004) proposes that we should treat belief as an "organic unity" of formation, sustenance, and belief-state. This helps make sense of how the value of a belief-state is influenced by its etiology. In particular, Zagzebski notes, the value of an organic unity of x + y can be greater than the sum of the values of x and y. She cites an example originally due to Brentano: pleasure is inherently good, while sorrow and wickedness are both inherently bad; but feeling sorry over wickedness is better than feeling pleased about it. Here, I will not address the question of whether beliefs really are unities that include their forming and sustaining processes, or whether epistemic value is determined by such unities as the value of a painting is determined in part by who painted it. For my limited purposes here, both alternatives are equivalent.

The credit theory of knowledge

The dominant virtue-theoretic account of the greater value of knowledge is the credit theory, due originally to Zagzebski (1996). On this view, when an agent knows, he has a belief that is true rather than false by virtue of his own endeavours; he "got things right owing to his own abilities, efforts, and actions, rather than owing to dumb luck, or blind chance, or something else" (Greco 2003b: 111). Thus to attribute knowledge to someone is to give them the credit for having a true rather than false belief.[9] To receive credit for a good outcome, one's success must not have been an accident. Thus the credit theory explains the sense in which knowledge is non-accidentally true belief (Riggs 2002).

More generally, however, a proper axiology for virtue theory should make a distinction between epistemic goods acquired accidentally, luckily, or haphazardly, and those acquired through the agent's own activities and capacities. A desirable property of belief has particular value if it obtains by virtue of the agent's doings. Call this general notion "subject-attributability", or SA; a property of belief is SA iff that property's obtaining is attributable to the subject rather than luck or some external feature.

The distinction here is intuitive; we regularly grant or withhold credit to agents for attaining certain goods themselves. Because of the recency of the idea, it has not been given a rigorous, satisfying development. In this thesis, we will make some progress towards a proper understanding of certain aspects of epistemic subject-attributability. All I need to establish here is that taking account of SA allows us to solve the problems adduced above. I propose that we take a property of belief to be valued both by its contribution to the goal of acquiring significant true beliefs and the extent to which the belief's having that property is attributable to the subject. Thus a belief's having property φ is more valuable if that belief's possession of φ is due to the subject's endeavours rather than luck or outside interference.

Praxical value

It is generally better to get something through one's own efforts—to achieve it oneself—than to be given it or to happen onto it. Earning money is

[9] In its present form, the credit theory appears to have trouble with testimonial knowledge. In at least most cases of knowledge by testimony, credit for true belief goes to the testifier rather than to the believer (see Lackey, forthcoming). It does seem, though, that to know the believer must at least receive credit for having chosen a reliable source. More work needs to be done on how the credit theory (and more generally, intellectual virtue) applies to secondhand knowledge. In the meantime, this thesis may be taken to deal exclusively with firsthand knowledge.

better than winning, finding, or being given an equivalent amount. Discovering the solution to a problem yourself is better than plagiarizing it. Making a good espresso yourself is better than buying one or blundering through the preparation. There seems to be a particular value in getting something because of a good performance, in "not just hitting the mark but hitting the mark somehow through means proper and skilful enough" (Sosa 2003c: 164).

Sosa calls the extra value arising from a good performance "praxical" value. It is a sort of instrumental value, in that what makes something a good performance (a good way to earn money, or solve a problem, or make an espresso) is largely that it tends to be successful. But it is also important that the tendency to succeed arise from the agent, rather than from happenstance. Praxical value is good in part because it ties the end result to the agent's efforts and abilities—it helps turn a good consequence into an achievement.

However, an unsuccessful act can still have substantial praxical value. A skilled but unlucky pull with a good machine is still better than an incompetent pull leading to an equally bad result. A student who finds a clever, but incorrect, solution to a problem still receives a fair grade; one who plagiarizes a wrong answer can't hope for much more than being the subject of an entertaining anecdote. This helps explain why it is better for a false belief to be reliably formed. An agent can pull off an admirable performance that is likely enough to be successful (and skilful) for the resulting belief to have value even though it was not in fact successful.

Indirect truth-conduciveness

In addition, goods achieved by one's own efforts are normally more readily preserved than those attained by other means or by luck. Thus, in attaining x oneself, one typically ends up with goods other than just x. Most obviously, the probability of future success conditional on attaining something oneself is increased. Having earned money yourself often implies being able to earn more money in the future. Only once in American history has a president who failed to win an uncontested plurality of the popular vote been re-elected. Likewise, having acquired a true belief through one's own abilities generally implies being able to acquire true beliefs in relevantly similar circumstances or on relevantly similar problems. Whatever abilities the subject had that allowed her to acquire a true belief in this case can generally be applied to similar problems or similar circumstances in future.[10]

One particularly important case of this is what we might call "temporal truth-tracking" (cp. Williamson 2000: 75-80). Our world is constantly changing in many respects, and thus successful beliefs often need to be regularly updated. Often, having a true belief at a single point in time is of little value to us; consider beliefs like "I will not fall through the ice", "I can get everything done before my appointment", or "I can hit that antelope with this rock." In such cases, what is important is that one be able to maintain true beliefs over time while one crosses the ice, one's appointment nears, or the antelope flees. For one of these beliefs to be accidentally true at t_1 says nothing about whether the agent will still believe truly at t_2. However, *knowing* any of these beliefs at t_1 generally implies being able to track the truth with respect to that proposition over time. This is another way in which attaining a true belief oneself is more valuable than just having the true belief.

In the *Meno* (97a-98a), Plato takes up the question of why knowledge is more valuable than mere true belief. He compares true belief to Daedalus's statues, which are so lifelike that they run away unless tethered to the ground. Knowledge, then, is like the tether that holds the statue in place. As the old saw runs, "easy come, easy go"; states attained by one's own powers are more stably possessed than those attained otherwise. For instance, the *New York Times* recently described how "financially lost winners [of lotteries are] the rule, not the

[10] In an unpublished manuscript, "Reliabilism and the Value of Knowledge", Alvin Goldman and Erik Olsson give a detailed analysis of the conditions under which the probability of future true beliefs, conditional on having a given reliably formed belief, is increased.

exception"; the author previously worked for a company that bought the rights to the future payments of lottery winners who had buried themselves in debt (Ugel 2007). Claudius assumed supreme authority in Rome by the sheer luck of being the only member of the imperial family to have survived Caligula's misrule, and it did not take long for his authority to be usurped by his wife, Messalina.

Similarly, suppose that to get to Larissa, I need to take the road south; suppose I believe that Larissa is north of here, but I also believe that I am heading north when I am actually heading south. I would be likely to run into evidence that would correct one of my two mistakes and lead me to go the wrong way. However, if I know where Larissa is, I am much less likely to replace my true belief with a false one (Williamson 2000: 78-9).

In both these ways, having knowledge is indirectly truth-conducive; the processes and environmental connections that generate knowledge tend to generate other true beliefs. Knowing that p may not satisfy our cognitive goals with respect to p any better than having an accidentally true belief that p, but it often leads to true beliefs at other times. Of course, new true beliefs are strictly speaking only epistemically valuable if they are significant. But if p is significant, it is likely that the extra beliefs that knowing that p can lead to are also significant. If p is significant at one time, it will usually still be significant at a slightly later time. It is also likely that if p is significant, the solutions to similar problems or beliefs involving similar propositional contents acquired in similar circumstances will also be significant.

Another axis of value

Recent work on the value problem has identified these general features of knowledge and observed that they help account for its greater value.[11] It does not

[11] On the first response, besides the Goldman and Olsson manuscript cited in note 10, there is Kristoffer Ahlström's "The Bearers and Makers of the Value of Knowledge". David Alexander's "Reliabilism, Epistemic Value, and the Normativity of Knowledge" examines the value of persistence. Patrick Rysiew's "Epistemic Agency and the Non-Local Truth Goal: A Defense of

appear, however, that we can entirely reduce the value of attaining something oneself to the extra goods thereby attained. Earning a fortune by designing slide rules or poodle skirts is better than winning it in the lottery, even though the former may not be any more indicative of future success than the last. Musicians are hardly better at keeping the money they earn than lottery winners. Recognizing the importance of subject-attributable epistemic goods allows us to make sense of the value of knowledge as an instance of more general forms of appraisal.

If we take subject-attributability to be another axis of epistemic value, we can get a richer account of our intuitions about the value of beliefs. In §2 above we noted that the products of some reliable processes are intuitively better than others. We can explain this by saying that sometimes a belief's being reliably formed can be attributed to the agent's own character and capacities rather than to something foreign or lucky. This allows us to make the intuitive distinction between perception on the one hand and reliable tumours or clairvoyance on the other. Subject-attributable reliable processes are of course virtues. Below, we will determine what virtues are in part by considering the conditions under which a belief's truth or reliability is attributable to the agent's own activities.

It may be objected here that the resulting framework is not an improvement on the muddles over justification that the ED framework is supposed to replace, because we have merely replaced one problematic concept, epistemic justification, with another, SA. But remember that the point of the new theory is not to make epistemological problems go away, but to provide a better framework in which to evaluate them. Much recent work can be understood as attempts to determine the conditions under which an epistemic property is attributable to a subject. Thus the effect should be the same as taking debates over justification to reflect differences as to which ED are of chief importance; it focuses the debate, allowing us to see what is really at issue, and *thereby* offering

Epistemic Value Monism", describes the general approach of accounting for the extra value of knowledge in terms of power (which he calls "non-local" truth-conduciveness).

more hope for a suitable resolution.

This is most obvious with knowledge. As I said above, recent work in virtue epistemology takes knowledge to be SA true belief. Greco (2003b) argues that we lack knowledge in Gettier and lottery cases because external fortuity prevents us from attributing credit for the true belief to the agent. Hawthorne's (2004) "moderate invariantism", according to which knowledge-claims are relative to the needs of the particular situation, can be seen as arguing that SA truth depends on considerations about the cost of being wrong in that case. And so forth.

A wide array of debates over conditions for knowledge and justification can be seen as attempts to elucidate appropriate conditions for SA. Thus, Plantinga's proper functionalism can be seen as the position that a property of belief resulting from the agent's functioning according to his design plan is thereby SA. Virtue theorists argue against this on the grounds that it pays insufficient attention to the subject's active role in inquiry. (The argument is most explicit in Zagzebski 1993.) Even classical internalism of the sort defended by BonJour (2003) can be seen as the position that a property is SA only if it is within direct, conscious access. We have of course already seen the problems with this view. I won't attempt to settle any of these issues; the most I can hope to accomplish here is to lay part of the groundwork for a proper understanding of subject-attributability.

Subject-attributability as epistemic value

One may wonder why subject-attributability is an *epistemic* value. In fact, virtue theorists sometimes wonder about this, and muse that the greater value of knowledge arises because "performances creditable to an agent as her own as the components of *eudaimonia*, of human good or faring well" (Sosa 2003c: 174). Some intellectual activity is surely part of *eudaimonia*, such as working on interesting and difficult problems, perceiving beautiful things, and the like. But it seems false that all the intellectual activity that leads to knowledge contributes to flourishing. It is hard to imagine how knowing that Casper is the capital of Wyoming or that there are exactly five dirty coffee mugs on my desk could make a greater contribution to *eudaimonia* than merely having the corresponding true beliefs.

The first reason for taking SA to be an epistemic value is that it appears to account for the greater epistemic value of knowledge; if subject-attributability were not epistemically valuable, than knowledge would not be more epistemically valuable than mere true belief. As well, as we just saw, we can understand many specifically epistemological theories as elucidating conditions for SA. So perhaps in §I.2 we misconstrued the cognitive goal; perhaps it is to not only have significant true beliefs, but to get them through our own efforts and capacities. Compare *athletic* values, which are determined by their contribution to achieving athletic success through one's own efforts. Winning by luck or by non-athletic means is not as valuable as winning through one's own skill and efforts.

It makes sense for acquiring significant truths ourselves to be the cognitive end when we remember that we are "finite knowers in a world we didn't make", as Quine said. Truths are not given to us, but need to be grasped; they are acquired, not found. We are not in a position to depend on fortuity; or more precisely, since there is always some luck involved in reaching a goal, there are limits to the extent that we can depend on external fortune. Our aim is thus significant truth and a measure of self-sufficiency.[12] (Or perhaps, as much self-sufficiency as we can muster, and the more the better; but for simplicity, I won't develop this possibility here.)

[12] To say that we wish to be self-sufficient doesn't mean that we are cognitive atoms. We depend on our epistemic communities for much of our knowledge, and we shouldn't suppose that our cognitive goal is in part to be free of this dependence.

As well, we might have to depend on having been designed in a manner that allows us to achieve our cognitive ends, or on the world's being governed by a loving God. Considerations like these would put bounds around the sphere in which our end is to be self-sufficient, without really altering the main thrust of the point. There are lots of directions in which this general point can be developed.

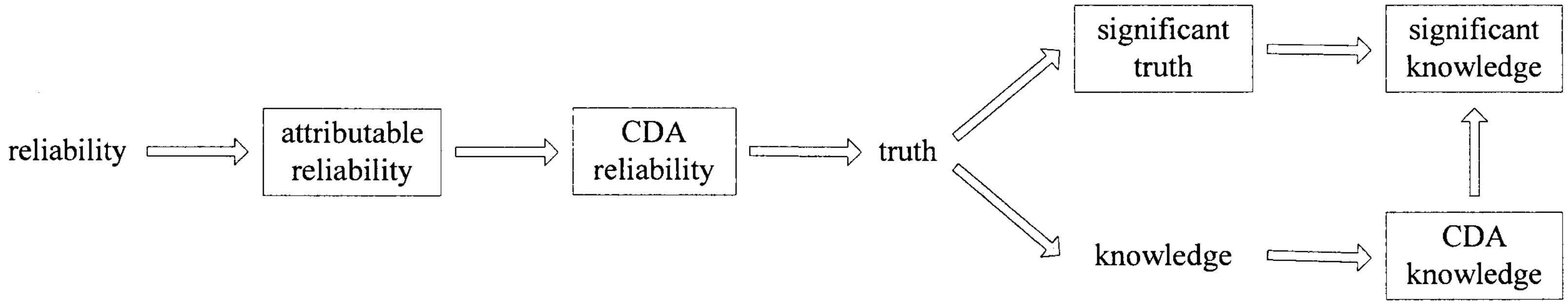
4. An epistemic desiderata framework for virtue epistemology

Even without a full analysis of SA, we can see roughly what effects it will have on the various groups of ED. SA true belief is knowledge, in accordance with the credit theory. When one attains something by one's own abilities, one is reliably successful across a range of nearby possible situations (Greco 2003b). Thus, for it to be SA that one has a true belief, one's belief must be reliably formed.

SA significant truth can be called "significant knowledge". The explanatoriness, coherence, and systematicity of beliefs appear to be more or less independent of whether they are achieved by the subject, and so they seem to have both SA and non-SA forms. But it would seem impossible to have understanding or wisdom without these goods being attributable to oneself; it might greatly advance our understanding of these values to determine how they are related to non-SA desiderata. SA reliability is, I'm suggesting, the better sort of reliability; and virtue theorists, proper functionalists, and the like have proposed different conditions for this ED. One would expect CDA to entail SA, since access to the grounds of one's beliefs is only valuable because it facilitates control over belief-formation.

A plausible framework for ED resulting from the inclusion of considerations of SA is given in figure 3. The arrows indicate greater value, with the same conventions and omissions as in figure 1. I assume in this figure that success in reaching our goals is more valuable, even when accidental, than reliability even when accompanied by SA. This seems to reflect the prominence of truth in common usage, though a more thorough investigation might indicate that this assumption is incorrect.

Figure 3. The revised ED framework



III

THE PLACE OF THE INTELLECTUAL VIRTUES

We ended the last chapter with the conclusion that epistemic desiderata are valued not just by their contribution to the end of significant true belief, but also by whether their attainment can be attributed to the subject's own doing. The need to understand subject-attributability is a primary motivation for investigating the notion of virtue, but not the only one. In this chapter, I will give a general account of the reasons behind moving to virtue epistemology—i.e., behind taking the intellectual virtues to be fundamental to epistemic appraisal. For this to make sense, of course, I will have to give a very sketchy, preliminary account of what intellectual virtues are supposed to be, and the tensions between different ways of understanding them. Then I will give a brief survey of the functions that intellectual virtues can serve in the theory of human cognition.

1. The aretaic orientation

The central idea of virtue epistemology is that neither reliabilism nor internalism can capture an appropriate sense in which agents are responsible for their own belief-formation. Reliabilist theories tend to treat agents as automata whose only contribution to intellection is having reliable processes. To understand our practices of epistemic evaluation, we need to acknowledge the difference between a reliable thermometer and an intelligent agent whose reliability is (in some sense) attributable to his own efforts and abilities. Creditworthy inquiry is an active, social process, and more than just having mechanisms that represent the world accurately.

Internalist theories are right to suppose that reliabilism cannot distinguish between correct inquiry and "flying blind" (BonJour 2003: 175), but they overestimate the importance of conscious reflection. As we saw with Stanley and Sally in §II.2, the status of belief sometimes depends on processes that are opaque to consciousness. In that case, it was because the process had been forgotten, but we will see more examples below.

The central proposal of virtue epistemology is that we can solve these problems and appropriately navigate between internalism and externalism by borrowing concepts from virtue ethics. In deontological and consequentialist ethics, as in traditional epistemology, the evaluation of acts is primary and the status of processes is derived from it. On utilitarianism, for instance, generosity is good because it tends to increase the overall happiness. On virtue ethics, the situation is reversed; the value of virtuous character traits is basic, and the value of acts is derived from the value of the traits that generate them.

Virtue epistemology started with Sosa's (1980) observation that reliabilist theories of knowledge were already fairly close to this inversion, since on those views the value of the belief is determined by how reliable its generating process is. But on simple reliabilism (as we have seen) it is unclear why agents can be praised or blamed for inaccessible desiderata. In virtue ethics, however, agents can be praised or blamed for acts that proceed from their character. Similarly, we could legitimately praise or blame agents for the properties of beliefs that proceed from their character traits. The result

> permits a motivated sensitivity to the complex interplay of internalist and externalist considerations in our practice of epistemic evaluation. This is because virtues are typically capacities, habits, or states of character that combine being internal to the agent with being such that their operations are largely opaque to reflection or introspection...Appeal to virtues offered a way of explaining how we are not alienated from our beliefs and inquiries in spite of the fact that we do not (and perhaps cannot) formulate or provide a non-circular vindication for the normative

standards which guide them (Hookway 2003: 184).

The intellectual virtues

Roughly, intellectual virtues are traits of character or cognitive dispositions that allow us to achieve our cognitive end of significant true belief. They are internal not in the Cartesian sense of being accessible to consciousness, but in the Aristotelian sense of being part of the agent's character.

There is a great deal of difference of opinion as to what the intellectual virtues are. It will help to give a few examples with the caveat that not all virtue theorists accept everything on this list. So here's what the intellectual virtues might include.

First, there are faculties that belong to persons, like perception, memory, rational intuition, and the like. Some of these might be learned or highly knowledge-dependent, like capacities for identifying birds or automobiles, or perhaps the capacity to understand a language.

Second, there are traits and skills that are important for successful inquiry and reasoning. Suppose one discovers that $\{p_1, \ldots, p_n\}$, all of which one believes, entail q, which is highly implausible. Given this, it is not necessarily a good idea to accept q; perhaps q is a reductio of $p_1 \wedge \ldots \wedge p_n$. The capacity to tell the difference between a reductio and an interesting consequence (or when to suspend belief in both antecedent and consequent) is a good candidate for an intellectual virtue. Consider too that the starting-points of any bit of reasoning have infinitely many consequences. Most of these will get you no closer to answering the question at hand, but some may be indispensable. The capacity to determine what implications are worth drawing out and examining is a plausible virtue. For a third example, consider the task of finding inconsistencies in one's beliefs. The general problem is in NP and presumably not humanly feasible; a collection of shortcuts that allows one to detect important inconsistencies is a possible virtue.

Third, we can identify what we might loosely call "intellectual personality traits". Some of these are intellectual versions of personality traits that are familiar in virtue ethics: courage, perseverance, curiosity, humility, open-mindedness, conscientiousness, impartiality. Others are unique to the intellectual domain, like originality.

Then there are some candidates that are more difficult to categorize. Most obvious here are *phronesis* or practical wisdom, and intelligence with its various subcategories. Such overarching capacities introduce complexities beyond the scope of the text, and I must therefore defer their discussion to another occasion.

Levels of intellectual virtue

The possible candidates for virtues are quite heterogeneous. We can make a significant distinction between the first category on the one hand and the latter two on the other. The first category contains skills and capacities for forming beliefs in certain ways. When praiseworthy (which all virtues are), they are reliable. Call these "low-level" virtues. The second two categories consist of traits that do not directly generate beliefs, but that

> regulate the ways in which we carry out such activities as inquiry and deliberation; they enable us to use our faculties, our skills, and our expertise well in pursuit of our cognitive goals (Hookway 2003: 187).

Call these "high-level" virtues.

There are *prima facie* important differences between these two categories and the sort of evaluations they are involved in. Low-level virtues can be cited as reasons for having knowledge, while high-level virtues cannot. One can say, "S knows that p because he deduced it from an obviety by obvious steps". But one cannot say, "S knows that p because he is original/conscientious/courageous," even though S might only have deduced that p because he had certain high-level virtues. It appears that the value of low-level virtues is transmitted directly to their products and only indirectly to the agents who have them, while the value of high-level virtues attaches directly to their possessor but only tenuously to their products.

Low-level and high-level virtues are in fact virtues in different senses of the Greek root *arete*. Low-level virtues are virtues in the sense of "excellence", which Sosa (1991: 271) and Greco (2000b) maintain is used by Plato and Aquinas to describe powers and faculties. Zagzebski, following the usage in Aristotle and in virtue ethics, maintains that "the Greeks identified virtues, not with the faculties themselves, but with the excellences of faculties" (Zagzebski 1996: 10). There isn't any real need to resolve this dispute; "intellectual virtue" is a technical term in epistemology (though one with a long history) and we are not beholden to prior usage. However, it does illustrate how "virtue" is used to describe two very different things—faculties, and capacities to use them well.

Another source of apparent heterogeneity in our list is that low-level virtues are all belief-forming capacities, while high-level virtues can sometimes be manifested in—or even by—the absence of belief. Consider someone defending a hypothesis against hasty rejection from others, suggesting a novel possibility for a solution to a problem, or listening to a student's challenge. These activities can manifest courage, originality, and humility respectively, but may not involve the formation of any beliefs. In fact, forming beliefs about the oft-rejected hypothesis, or the novel possibility, or the student's challenge might manifest vices of excess.

Virtuous belief-formation, of course, involves avoiding error as well as believing truths. It is quite likely that any case where suspension of belief or adoption of a non-doxastic state is virtuous is a case where one is likely to fall into error by forming a belief either way. So we can still understand the value of these virtues as reducing to their contribution to the acquisition of significant truths. This is, after all, the ultimate goal, and indecision merely its prelude.

Character traits can contribute to our cognitive goals in different ways.

Being reliable is one, but this comes in degrees. A process can be reliable to a certain extent, but not (say) reliable enough for us to say that its products are known. A process or trait can also be truth-conducive by broadening the range of true beliefs that can be formed (which Goldman calls *power*) or the range of environments in which true beliefs can be formed (*portability*; see §IX.1 for discussion). A process can also be particularly valuable by virtue of generating significant beliefs. This seems to be where much of the value of, say, originality comes from. Original thinkers are probably no more *reliable* than the rest of us, but they generate significant beliefs that we cannot otherwise come by, and thus make a special contribution to the community's intellectual development. (See Zagzebski 1996: 182-3.)

So there are plenty of ways that virtues can contribute to the goal of acquiring significant true beliefs. (They will be discussed in more detail in ch. IX.) It is thus plausible that the high-level virtues are such in part by virtue of contributing to our cognitive goals. Moreover, the concept of virtue in both the general sense of "excellence" and the sense in which it is used in moral theory implies that one will usually be successful in achieving one's aims (see Zagzebski 1996: 176-86).

2. The role of virtues in epistemology

Having seen a brief account of the intellectual virtues, we can turn to summarizing some of the epistemological problems that it appears they can help us solve.

Knowledge and other classical problems

We have already seen one role the virtues can play in epistemology—helping us understand when the achievement of an epistemic desideratum is attributable to the agent herself. Thus it is quite plausible that virtues are involved in the analysis of knowledge. Virtues can also be seen as central to directly truth-conducive desiderata. As we saw, not all reliably produced beliefs have the same intuitive status. It is plausible that the difference between a reliable tumour and vision is that the latter is, and the former isn't, a virtue.

Virtue theories are particularly valuable for understanding how we evaluate beliefs that are not supported by reflectively available evidence. As Greco (2001) argues, they can explain how knowledge can be produced by processes that are not regulated by propositional-level rules, such as connectionist networks that might correlate inputs and outputs at the propositional level but that only implement rules at a subrepresentational level.

Intellectual virtues play an important role in responses to skepticism. For instance, Greco argues that inductive skepticism arises because inductive inferences are only contingently reliable. To explain, then, how an agent can legitimately make inductive inferences, we need to explain how she can be appropriately sensitive to a contingent, non-logical evidential relation. This, Greco argues, can only be appropriately explained by her having character traits that reliably make inductive inferences, and that she uses conscientiously (2000 chh. 6-7).

Greco's approach uses virtue theory to make externalist responses to the problem of induction (e.g., van Cleve 1984) more plausible. Hookway takes a more radical view of skepticism. He argues that skepticism arises because we have to reflect on our activities in inquiry and deliberation, but there is no clear limit to how much reflection is necessary. Thus skepticism "places upon us burdens of reflection that we cannot discharge" (2003: 197). The habits and tendencies that allow us to avoid being trapped in indecision must be virtues. Only by being "opaque to reflection" (198) can they actually block the need for an infinite regress of reasons, but they must also proceed from one's own character so that their influences are not felt as "alien" or "heteronomous" (199) and thus demand reflection before being acceptable.

The virtues and nonclassical epistemic evaluations

Obviously, virtuous character traits are central to our appraisals of agents themselves. These can go beyond the ED of the agents' beliefs. Suppose, for instance, that A teaches B everything he discovers. Any desideratum that A has, B also has. But since B gets all these desiderata from A, A is the superior epistemic agent. (See Zagzebski 1996: 26-7.) The difference cannot be cashed out in terms of reliability or other ED, but it is plausible that A possesses virtues that B lacks.

Our choices of intellectual exemplars are similarly divorced from the possession of epistemic desiderata. Montmarquet (1993) notes that thinkers such as Aristotle, Newton, and Einstein are all apparently more or less equivalent in intellectual virtue, but the quality of their beliefs differs enormously. More generally, we do not appear to appraise past agents just on the quality of their beliefs. It does not take that much education in modern science to reach a point where the quality and quantity of one's beliefs about the natural world exceeds Aristotle's; the same can be done, though with more difficulty, for Newton. But, of course, vanishingly few people are their equals in intellectual virtue. This problem will be taken up again in ch. IX, where we will see how the concept of virtue allows us to understand such evaluations.

The virtues are often thought to play a special role in ED other than truth, knowledge, and justification's descendents. For instance, Riggs (2003b) argues that the virtues are partially constitutive of wisdom. Our task here is not to analyze cognitive goals such as wisdom and understanding. We should recognize, however, that a virtue may be especially valuable because of the significance of the beliefs it generates. This gives us room to extend our virtue theory in another place by spelling out the sort of significance that virtues involved in wisdom and understanding produce.

We should note here that virtues can contribute to the attainment of significant truth in several ways. One is to avoid error; i.e., to be reliable or improve the reliability of one's beliefs. Another is to increase the range of true beliefs one can form, thus combating ignorance. Goldman (1986: 26-7) calls this *power*. A third way is to extend the range of environments in which one can meet one's cognitive goals or increase the speed with which one can make decisions. Call this *portability*; it and power will be discussed further in §IX.1.

The two aspects of the goal of attaining significant true belief—attaining significant truth and avoiding error—can sometimes conflict (see Riggs 2003a). One might speculate that virtues are involved in finding an appropriate balance between these two goals. Thus virtues may help us determine how to weight epistemic desiderata and what ED to aim for in different situations.

In this thesis, I am only be concerned with individual cognition. Intellectual virtues are likely to be important for understanding group cognition as well—research teams, group problem-solving, and the like. This is partly because of the overlap between the intellectual and moral virtues. A virtue like courage that is important to both social and intellectual life is presumably particularly valuable in a group setting. We can also speak of the virtues of scientific theories, and so the virtues might have an important role to play in philosophy of science (Hookway 2003).

Virtues in decision theory

More generally, much of what we use in decision-making must be "opaque to reflection". Virtues can play an important role in understanding how our inquiries can be appropriate and responsible despite not being supported by explicit, introspectible reasons. Human cognition is subject to onerous limits on both computational resources and on the time available for deliberation. Many of the problems that rational agents need to solve—for instance, finding decisions that maximize expected utility—are effectively intractable given the available computational and temporal resources. To make problem-solving tractable, embodied agents need to make use of heuristics and shortcuts.[1]

It appears that virtues have a special role to play in evaluating techniques for managing resource limitations. Morton (2004) notes that given almost any distinction between problems too hard to solve and problems that are tractable, the problem of identifying whether a problem is too hard to solve will sometimes itself be too hard to solve. Thus the use of heuristics can never be entirely justified by explicit reasoning or canonical decision-making methods. For instance, it is often argued that satisficing—choosing the first option with satisfactory utility, instead of evaluating all options and picking the best one—is a way of finding an optimal balance between payoffs and decision costs. Satisficing is optimal when the utility of choosing the first satisfactory option minus the (relatively low) costs of making that decision is greater than the utility of choosing the option with the greatest utility minus the relatively high costs of enumerating all the options (Byron 1998). It's not hard to see that the problem of determining whether one has attained the optimal balance between utility of outcomes and decision-making costs is just as complex as the problem of finding the option with the highest utility.

Virtues can provide a way of evaluating appropriate decision-making despite the lack of explicit reasons for its rationality. Taking a particular shortcut out of virtue is not the same as guessing that that shortcut will work, although neither is supported by sufficient introspectible reasons. Moreover, encouraging the development of virtues might very well be an effective way to improve human decision-making when presented with problems that are intractable using more canonical methods. And besides evaluating how we perform these tasks, virtues can also play an explanatory or descriptive role in understanding how decisions are made under constraints (see, e.g., Morton 2003: ch. 2). This use of virtues is linked to the distinction we saw in ch. II between virtues and reliable processes

[1] For the general approach, see Simon 1982; see Gigerenzer et al. 1999 for more detailed analyses.

that do not (intuitively) yield knowledge. Adaptability and flexibility are hallmarks of rational behaviour; virtues can help us understand how this is possible even in decision-making not based on explicit reasons and canonical ratiocination.

There are two approaches to the involvement of intellectual virtues in rational decision-making (Hookway 2003). On the one hand, we might suppose that virtues allow us to apply heuristics and solve problems that we otherwise could not solve, but are only instruments for embodied agents to make rational decisions (which is approximately Byron's 1998 position). On the other hand, we could take the manifestation of the virtues to be constitutive of at least some forms of rationality (cp. Swanton 1993). Analyzing rationality is not my aim here, and I will thus be evaluating the contribution of the virtues to proper decision-making in a purely epistemological (and therefore somewhat narrow) manner. Thus they are instrumental in generating correct decisions inasmuch as they contribute to the acquisition of true beliefs of the form "I should do x". Production by the virtues can be constitutive of a correct decision's being attributable to the agent (rather than, say, to a lucky guess).

As well, one can be committed to the virtues' being central to inquiry and decision-making, but not to their being involved in knowledge or other epistemic desiderata. I'll call this position "non-epistemic virtue theory". One form of it (Morton's) will be discussed in the next chapter.

Exploratory capacities

Humans are inherently exploratory creatures; we have a basic drive to venture into and map the unknown. This exploration can be literal, when we physically enter new regions, or figurative, as when one learns a new subject or takes on a new job. Moreover, the capacity for successful exploration is vital to our overall success in life.

Exploratory creatures cannot just depend on environmental regularities to give them reliable beliefs, because these regularities may not obtain in new environments. A person living in Arcadia, where everyone is simple and honest, can reliably acquire true beliefs just by trusting whatever anyone tells them. But this capacity won't give her the ability to get by in Crete. It is plausible that exploratory capacities must be grounded in the agent's own character; there is not much besides your character that you take wherever you go.

Moreover, capacities to adapt to new circumstances (and thus to explore new regions) can go beyond what is provided by explicit reasons. Take finding a new scientific paradigm, for instance. Kuhn (1970) observes that scientific revolutions typically involve revising canons of evidence and introducing new concepts for describing phenomena. As a result, the new theory is typically neither expressible in the vocabulary of the old one, nor justified by the prior evidential standards. Peterson (1999) argues that this is only a special case of a ubiquitous phenomenon. It is typically easier to know that one is doing something wrong than it is to know what the right thing to do is. Adapting to new situations often involves adopting new concepts and new beliefs for which there are insufficient reasons accessible by the old way of doing things. Saving a failing marriage, for instance, often involves learning new ways of interpreting events and revising one's beliefs about what one should do or infer. These new techniques are typically underdetermined by reasons accessible from the old way of looking at the situation. For these reasons, it seems potentially fruitful to treat adaptive capacities as virtues.

The union of ethics and epistemology

Finally, virtue epistemology might allow a unification of sorts of ethics and epistemology. There is a *prima facie* value to having a more or less unified account of human normative theory. Some theorists (most notably, Zagzebski 1996: 137-64) have proposed that there is no essential difference between moral and intellectual virtues. However, there appear to be cases where the moral and intellectual virtues conflict. Driver (2000), for instance, argues that good and loving parents should have an exaggerated view of their children's accomplishments, since honest praise is important for a child's success. That would seem to indicate that the moral virtues of good parenting are incompatible with the epistemic virtues of dispassionately believing the truth.

The possibility of unifying ethics and epistemology depends on the supposed analogies between the two fields that we saw in §I.1 to be problematic because of the differences between the types of entities being evaluated. I will not assume that any aspect of moral virtues can automatically be carried over to a theory of intellectual virtues. All proposals, regardless of their source, will be judged exclusively on grounds of their suitability to epistemology. Perhaps at the end of the day there will turn out to be extensive commonalities between the two fields. Or not; it's not a matter that I will address here.

This is a long wish list, and one might expect that as the concept of virtue is made more rigorous it will turn out not quite to satisfy all these roles. However, as we develop the notion we should bear in mind the various uses to which it might be put, and try to develop it in a way that will allow it to have the widest possible explanatory role.

IV

THEORIES OF VIRTUE

In the last chapter, we saw that virtues are, very roughly, traits of character that allow us to achieve our cognitive goals; we saw some examples of putative virtues and some of the roles that virtues might play in epistemology. In this chapter, we'll look at some theories of virtue.

Virtue theories are generally divided into two types: *virtue reliabilism* and *virtue responsibilism.* The former are (as the reader might guess) close to reliabilism; the latter are inspired more by virtue ethics. For purposes of exposition here, I'll add a third category of virtue theory, which I'll call *non-epistemic.* These are theories of how the virtues figure in reasoning and decision-making. They are importantly different from other virtue theories because they do not imply that the virtues are involved in the production of knowledge. Thus they are worth treating separately, even though the difference is more in the approach to virtue than the view of virtue itself; non-epistemic theories come in reliabilist and responsibilist flavours.

We will see below that all of these approaches tend to favour one sort of virtue over others. Virtue reliabilism handles low-level virtues well, and has trouble with high-level ones. Virtue responsibilism and non-epistemic theories are the opposite. This might make it plausible that there is a fundamental distinction between low-level and high-level virtues, and we could, say, have a reliabilist view of the low-level virtues and a responsibilist view of the high-level ones.

Alston describes the attempt to explain all or most epistemic desiderata in terms of virtues as driven by "imperialist pretensions" (2005: 3), and so we may as well call the approach I'll take in this thesis "methodological imperialism". I will try to find an account of virtue that satisfies as many as possible of the putative roles for virtue in epistemic evaluation that we examined in the last chapter.

1. Virtue reliabilism

Virtue reliabilism is most prominently espoused by Ernest Sosa and John Greco. Alvin Goldman has defended a virtue theory along similar lines, which I will also discuss.

The central idea of virtue reliabilism is that virtues are stable dispositions to form true beliefs that are part of one's cognitive character. We will take up what constitutes cognitive character in detail in the next chapter, so here I will only give a brief summary of the important points. Sosa takes virtues to be part of or derived from one's "inner nature" (1991: 140); Greco takes them to be abilities to form true beliefs in certain ways (2000, 2003b). They must be counterfactually stable in the sense of obtaining over a range of nearby possible worlds, provided that the agent's character is not unduly altered. (So, for instance, if S pokes his head above the trench and narrowly avoids being shot, he still has an ability to see the enemy's lines even though he could easily have been dead.) Virtues need not, however, be temporally stable (Greco 2003a: 470-1), and may be either innate or acquired (Sosa 1991: 277). In both these respects they differ from Aristotelian moral virtues, which must be temporally stable and acquired by habituation. Agents are held responsible for having or lacking virtues not in a moral sense, but in the same sense in which they are responsible for having or lacking other abilities (Greco 2000a: 211-7).

The approach is similar to Plantinga and Millikan's proper functionalism.[1] On that way, a process can only generate knowledge if it was designed well and is

[1] See Sosa 1993, in which he calls Plantinga's theory a form of virtue epistemology, and Plantinga's (1993c) response, in which he calls Sosa's theory a form of proper functionalism.

performing its intended function in the intended circumstances (Plantinga 1993a, Millikan 1993). Either God or natural selection may be responsible for the design of our faculties. The two views treat knowledge in very similar ways, but there is a deeper divergence in the ways they conceive of praiseworthy epistemic behaviour. As we saw in the last chapter, virtue is a matter of how the agent is and what he does; virtue proceeds from one's own character. Proper functions are dictated by one's design plan, and thereby proceed from something external to the agent's actual constitution and activity. Thus proper functions can't accommodate the broader project of virtue epistemology (see Zagzebski 1993 and Axtell 2006).

Let us now turn to the specifics of the three dominant reliabilist virtue theories: Sosa's virtue perspectivism, Greco's agent reliabilism, and Goldman's virtue theory.

Sosa's virtues

Sosa combines virtue theory with perspectival internalism. A virtue V(C, F) is a disposition to believe correctly on propositions in F in conditions in C that is derived from the agent's inner nature (1991: 138-42). (We'll look at where the C, F pairs come from below.) Agents also cannot be disposed to overapply their faculties in conditions in which they are unreliable. This condition is meant to handle cases like Mr. Magoo, an old-time cartoon character who is extremely nearsighted but forms mistaken visual beliefs anyways, to comic effect. Mr. Magoo has a disposition to form true visual beliefs at close distances, but fails to have a virtue because he overapplies his visual processes.

Agents always have virtues relative to an environment E, which Sosa regards as a stable background, broadly construed (284). This is meant to explain our intuitions regarding the epistemic status of brains in vats or victims of Cartesian evil demons (which Sosa calls the "new evil demon problem"). These are imaginary duplicates of ourselves with exactly our beliefs and experiences who live in illusory worlds and whose faculties are thus (so says Sosa) completely unreliable. However, there is a strong intuition that there is a sense in which brains in vats that are duplicates of praiseworthy real-world agents are "justified" despite their unreliability. Sosa's proposal is that since they form beliefs the same way we do, they have virtues relative to the actual world, but not relative to their own worlds. When we appraise agents, we usually consider whether they have virtues relative to the surface of the actual Earth, even when they are only possible agents who would live in very different situations (144). Thus we are inclined to regard brains in vats as virtuous.

This response requires taking the conditions C in which one has virtues to be defined by properties that can obtain both in illusory situations and the actual world (or by a disjunction of subjectively indistinguishable illusory-world and actual-world conditions). Below, when we examine Sosa's response to the generality problem, we will see that he individuates virtues primarily from their possessor's point of view.

Goldman has objected that there is insufficient reason to conclude that in ordinary practice, we relativize epistemic evaluations to possible worlds (1992: 161-2). He proposes instead that ordinary epistemic evaluations are made with respect to the way things are in the actual world. It would be sufficient for Sosa's response to the new evil demon problem to suppose that we always evaluate virtues with respect to the actual world, or at least to worlds relatively similar to our own (so that worlds in which we were brains in vats would be excluded). This issue is quite difficult to adjudicate, and concerns details that we needn't worry about here, so let us move on to virtuously held belief.

Sosa calls a belief produced by an intellectual virtue operating in C and F for which it is reliable an *apt* belief (1991: 289).[2] Aptness is a sort of justification (2003a), although it does not entail that agents have consciously accessible reasons for believing. It is nonetheless the sense in which the foundations of

[2] More recently, he calls a belief produced in the right circumstances by a faculty that is a virtue relative to the actual world 'adroit' (2003a: 157).

knowledge are justified. Apt true belief is *animal knowledge*, which is meant to be the commonplace, unreflective knowledge that we have so much of.

Virtue perspectivism

Sosa takes BonJour's examples of reliable clairvoyants to show that there is more to knowledge than animal knowledge. Reliable clairvoyants have virtues, but lack accessible reasons for their beliefs. Similarly, there seems to be a sense in which brains in vats have the same reasons for their beliefs as us, since they have the same experiences and believe they have the same faculties.

So Sosa proposes that there is another sense of "justification" that is a matter of having an appropriate perspective on one's beliefs. The meta-beliefs making up this perspective must (a) indicate the source of one's beliefs and that it is virtuous; (b) provide appropriate inferential and evidential connections that make one's belief-set into a coherent whole.

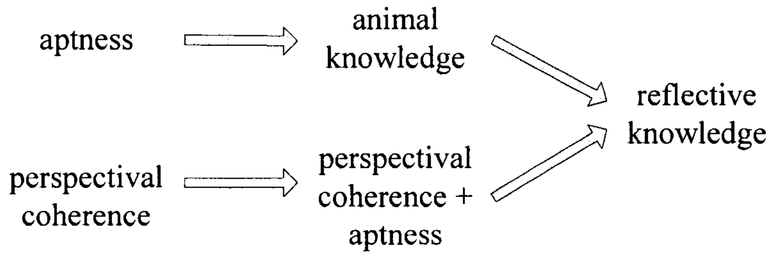
Meta-beliefs arise from "coherence-seeking reason", i.e., the process of maximizing the coherence of one's belief-set. When one's object-level beliefs are mostly apt, coherence-seeking reason is a virtue, since it is reliable (in the actual world) when applied to sets of mostly true beliefs (Sosa 1995). Thus one's meta-beliefs are apt.

This is a very coarse account of what agents have to do when trying to increase the coherence of their beliefs; "coherence-seeking reason" runs together a large number of potentially disparate processes and tasks into a single grand faculty. This is especially problematic because it obscures the way that different high-level virtues alter our beliefs in different ways.

Suppose, for example, you are trying to incorporate together a number of approximately equally plausible beliefs that contradict each other. Sometimes you can find a way to fit them all together into a single coherent theory; but this is the stuff of which genius is made, and you can't count on it. That leaves a variety of approaches lying between two poles. One possibility is to pick a consistent subset and build a coherent theory out of them. This maximizes coherence, but in the absence of militating reasons for keeping one belief over another, you can't be sure that the ones you retained are the true ones. At the other extreme, you can maintain them all and learn to live with a certain amount of inconsistency (for instance, by restricting the domains in which you draw conclusions from different principles, as is done is contemporary physics). The latter strategy is safe, though it means accepting a demonstrably false conjunction of propositions; the former avoids that problem, but at the risk of inadvertently choosing consistency over truth. These are very different sorts of approaches, and manifest different sorts of intellectual virtues. (Cp. Russell 1946: 592, where he maintains that the use of the latter strategy is a virtue of Locke's that other historical philosophers lack.)

Nonetheless, as long as the processes one uses to generate a coherent belief-set are virtuous, then the essentials of Sosa's theory are untouched. Apt, perspectivally justified true belief is *reflective knowledge*. Reflective knowledge is a higher epistemic goal than animal knowledge. It is closely linked to understanding (cp. Grimm 2001) and is the Cartesian ideal (Sosa 1997). The resulting system of ED is found in figure 4. Note that Sosa may acknowledge ED (like wisdom) that are not included in the figure, as it includes only those he discusses. I assume that aptness in the figure is relative to one environment E. Sosa gives no indication as to whether aptness relative to certain environments is

Figure 4. Sosa's ED framework



more valuable than others. One might take our disposition to evaluate virtues relative to the actual world to indicate that actual-world aptness is more epistemically valuable, but there are also reasons to think that aptness relative to one's own environment would be more important to have.

Sosa (1991: ch. 11) argues that there are two routes that lead to virtue perspectivism. The first is the one we've followed; to start with simple reliabilism and recognize that there are ED that involve coherence and having accessible reasons for beliefs. The other is to start with a coherence theory and recognize that coherence is only truth-conducive if reliable faculties anchor the belief-set to the world.

The generality problem

The epistemic perspective is also important for handling the *generality problem*.[3] This is a worry for any position on which reliability is, or is a necessary condition of, an epistemic desideratum. I will discuss the generality problem for reliabilist theories overall in ch. VIII; here, let us just see how it applies to Sosa's views specifically. Any belief is formed in a token environment and is a belief in a single proposition. However, reliability on Sosa's account is a matter of performance over a field of propositions F in a range of conditions C. So the problem for Sosa is identifying appropriate C and F for which it is meaningful to call a process a virtue. These must be neither too narrow nor too broad. For instance, if S forms a true belief that p in situation s, then S has a virtue V(s, p) regardless of how irresponsibly he came about the belief that p. This is the "single-case problem". Similarly, if F contains only necessary truths, then every agent capable of believing those truths has, for any C, a virtue V(C, F). At the other extreme, we face the "no-distinction problem". Any beliefs that arise from the same virtue are equally reliable; if C and F are too broad we run the risk of lumping together beliefs with intuitively different statuses.

[3] See Conee & Feldman 1998 for the canonical statement of the problem.

Sosa takes attributions of reliability to be important because of our need to identify trustworthy informants, which can be either other agents or our own faculties. For our own faculties, he argues, an accidentally true belief is one that is accidental relative to our perspectival understanding of our faculties, where this exists (1991: 282-3). Our meta-beliefs will tend to lump together certain (C, F) pairs into groups that are all treated alike, and the resulting pairs are the dispositions that we can evaluate as being virtues or not. For agents that lack detailed perspectives, Sosa takes habits of belief-formation and implicit expectations to do the same job (2004: 294-7).

Sosa also maintains that the virtues of agents without perspectives can be individuated by their epistemic community's needs to use them as reliable informants. This means that (C, F) pairs must be projectible and likely enough to occur at different times to be worth picking out (1991: 281-2). This makes the possession of virtue dependent on the surrounding community, rather than on the agent's own cognitive system, which is a very peculiar position to take on virtue. It also threatens to make attributions of animal knowledge "metaphorical" (275). Thus his more recent work on individuating virtues focuses on the implicit expectations formed by belief-forming habits.

Agent reliabilism

Greco argues that neither of Sosa's two kinds of knowledge properly captures the concept; the conditions for reflective knowledge are too strong and those for animal knowledge are too weak. Very few knowers have the detailed, coherent perspective that is necessary for reflective knowledge (2000a: 188-90 and 2004). At the same time, animal knowledge only guarantees that beliefs are formed in an externally reliable manner. Virtues must be stable dispositions of subjects, which rules out strange and fleeting processes, reliable brain tumours, and the like. But "knowledge has to be subjectively appropriate as well as objectively reliable" (2000a 180), and animal knowledge does not capture this sort of subjective justification.

This is illustrated by reliable clairvoyants. Being a disposition to believe correctly in certain circumstances, clairvoyance is a virtue for Sosa as long as it arises from the clairvoyant's inner nature (which is easy to imagine without altering the thought-experiment). Suppose Norman believes that the President is in New York because of his clairvoyance, without supporting evidence, while Orson has the same belief because he has just met the President there. The only difference between Norman and Orson, on Sosa's account, can be in their epistemic perspectives. Greco argues that if Orson is a typical sort of unreflective person, it is implausible to suppose that he has an epistemic perspective sufficiently detailed and coherent to yield reflective knowledge. So we must capture the difference between the two in some other way.

Greco develops a very general position he calls "agent reliabilism", which he takes to capture the essence of virtue perspectivism (as well as several other authors' positions; see 2000a: 178-9). On agent reliabilism, a virtue is an ability to believe correctly that arises from the agent's cognitive character. It must be a stable disposition, and must be integrated appropriately with the rest of the agent's character. Greco (2004) endorses Sosa's attempt to solve the generality problem in terms of the surrounding community's needs for informants (which as I noted above, is problematic), but does not develop a full response to the generality problem.

Nor does he develop the notion of cognitive integration very much. It is "a function of cooperation and interaction, or cooperative interaction, with other aspects of the cognitive system" (2003a: 474). This makes it at least a matter of the range of beliefs generated by the disposition, the extent to which outputs of the disposition are related to other beliefs instead of peripheral to the belief-set, and the sensitivity of the disposition to defeating evidence.

Still, cognitive integration doesn't capture the sense in which knowledge is subjectively justified, or subjectively virtuous. That is a matter of using dispositions that one countenances, i.e., that one generally manifests when thinking conscientiously. One thinks conscientiously when one's sole motivation is to get the truth; it is a default from which vicious motivations disturb us (2000a: 179). A belief arising from a countenanced disposition is well formed from one's own point of view because it arises from the stable aspects of character that one generally applies when motivated to believe the truth. Agents are sensitive to the grounds of beliefs arising from countenanced dispositions not in the sense of being able to consciously access those grounds, or having a perspective upon them, but in the sense that forming those beliefs from those grounds is something they would do when on their best behaviour (so to speak).

Of course, one might be wrong about when one is using countenanced dispositions, just as in general one can be wrong about one's motivations. A father might try very hard to disinterestedly evaluate the evidence regarding his son's crimes but unknowingly believe out of dispositions that he would not countenance (191). Nonetheless, Greco argues that since we are often not aware of the etiologies of our beliefs, we are not guaranteed to be aware of their subjective status for us (see Greco 1990). I have heard that it is common for subjects to believe that Coke tastes better than Pepsi, even though in blind taste-tests they prefer the latter; it appears that they prefer Coke because of the cachet it acquires through its advertising, and misattribute their preference to flavour. The belief that Coke tastes better than Pepsi is ill-formed from most persons' own point of view, but being unaware of this, they detect no reason not to continue with the belief.

An agent who has knowledge uses a countenanced disposition that is also a virtue (and thereby acquires a true belief). In knowledge, "reliability results from responsibility" (1993: 429); the agent has a reliably formed belief because she reasons conscientiously. Her reliable character is moreover the most salient explanation of her having a true belief, and thus she receives credit for it (2003b).

The combination of virtue and subjective justification allows Greco to explain the problem with reliable clairvoyants. He argues that if Norman's cognitive system is like ours with clairvoyant powers added, then either (a) Norman does not countenance his clairvoyance or (b) Norman's clairvoyance is insufficiently integrated with the rest of his cognitive system—particularly his system for handling defeating evidence—to be a virtue. Greco argues, however, that if Norman's defeater system is quite different from ours and he countenances his clairvoyance, then it is possible for it to generate knowledge for him. What we have in that case is just a person with an odd cognitive system that gives him a very unusual virtue (2003a 474-6). This seems to be entirely on the right track (and I'll argue for a very similar reading of the case in §VII.1), although a great deal here depends on the underdeveloped notion of cognitive integration.

Goldman's virtue theory

Goldman (1992) proposes a very different sort of virtue theory that is sometimes regarded as a form of virtue reliabilism. He distinguishes between the task of describing our commonsense epistemic evaluations and the normative task of determining how we ought to form beliefs. His virtue theory is an attempt to explain how we make epistemic evaluations. He proposes that when we evaluate the justification for a belief, we compare its etiology to a stored list of virtues and vices. The belief is justified if its etiology is sufficiently similar to a virtue; unjustified, if sufficiently similar to a vice; and neither if its etiology does not match a process on either list. "Justification" here is very closely linked to knowledge. We advert to virtues when answering the question "how does x know?", and barring Gettier complications, justified true belief is knowledge. As a result, this gives only an account of low-level virtues; as we saw in the last chapter, one cannot advert to high-level virtues in explaining how someone knows.

Processes are classified as virtues if they appear to be reliable in the actual world and as vices if they appear unreliable. The resulting lists are largely learned from the community, and reflect longstanding practice more than individual judgments. We do not revise our lists of virtues and vices in hypothetical scenarios; when considering other possible agents, we evaluate not whether their beliefs would be reliably formed, but whether they use processes that would be reliable in the actual world.

This conservatism allows for solutions to the usual counterexamples to reliabilist theories. Brains in vats are as justified as we are because they use the same virtues that we do. Reliable brain tumours resemble "pathological processes" (1992: 159), which are vices, and thus the beliefs they produce are unjustified. Clairvoyance does not match any virtue on the list, but probably does not match any vice either; thus beliefs produced by it are neither justified nor unjustified.

Thus virtue is in the eye of the beholder, and there is no inherent link between virtue and cognitive character. In this sense, Goldman's position diverges remarkably from other virtue theories. This becomes more apparent when we turn from the project of describing how we make epistemic evaluations to the normative project of revising our practices of belief-formation. Goldman conceives of epistemology as continuous with cognitive science. Scientific investigation allows us to acquire a more thorough and careful understanding of how we form our beliefs and how our processes contribute to our goals of acquiring significant true belief, which allows us to make more precise and accurate epistemic evaluations.

However, virtues and vices do not appear to play any role in normative epistemology, except when we update our lists of virtues and vices as a result of new information about our processes. Goldman's normative epistemology does not consider whether the attainment of desiderata is attributable to subjects or not. It is also problematic that this position only accounts for low-level virtues, and entirely unclear how it could be extended to cover high-level virtues as well. So on the conception of intellectual virtues we are working with here—truth-conducive traits of cognitive character—Goldman's theory is a form of process reliabilism augmented with an explanation of why epistemic evaluation deviates from what reliabilism recommends.

2. Virtue responsibilism

Whereas virtue reliabilism sticks quite closely to other forms of reliabilism, virtue responsibilism is based on the structure of virtue theories in ethics. The position is most prominently espoused by Linda Zagzebski, Guy Axtell, Abrol Fairweather, and James Montmarquet. Lorraine Code's (1987) virtue theory is an important forerunner. The central idea of the approach is that just as the moral virtues are tied to responsible behaviour, the intellectual virtues are tied to responsible believing.

Both types of virtues have essentially the structure of Aristotelian virtues. These are much like skills. But whereas skill is identified by the capacity to reliably produce certain consequences,

> if the acts that are in accordance with the virtues have themselves a certain character it does not follow that they are done justly or temperately [for instance]. The agent must also be in a certain condition when he does them: in the first place he must have knowledge, secondly he must choose the acts, and choose them for their own sakes, and thirdly his action must proceed from a firm and unchangeable character (NE, ca. 1105a30).

There are certain glosses to these conditions that appear to be standard among Aristotle's followers. Virtue requires both capacities for accomplishing certain ends and the ability to know when it is right to apply them. A benevolent person needs to know not just how to help others, but also when it is right to do so. A person is a skilled arms dealer because she can sell arms effectively, not because she knows when it is right and when it is wrong to do so. Virtues and vices are acquired by habituation, and cannot be innate. This is part of why we are responsible for our virtues and vices, though we are not necessarily responsible for innate traits. Virtuous acts must be motivated to achieve their proper end. A student who works in a soup kitchen in order to increase her chances of getting into Harvard is not acting benevolently; a truly benevolent person is motivated to help others for *their* sakes. Virtues must also be reliably successful in attaining the end by which the agent is motivated, though a virtuous act need not succeed on any particular occasion. Finally, virtues are robust character traits. Once acquired, they are not readily lost (or cannot be lost at all). Moreover, they are manifested in every, or nearly every, situation that calls for them. The benevolent person is benevolent all the time (or just about).

This is the general picture. It shouldn't be hard to see that this general approach is going to have trouble with low-level virtues. However, that's a matter I will address in detail in §V.1. Zagzebski has provided the most systematic and rigorous theory of virtue responsibilism, and so I will now examine her views in detail.

Zagzebski's responsibilism

Zagzebski's virtue theory applies to both moral and intellectual virtues. She does not think there is any significant difference between the two categories, and virtue-words that refer to traits in both fields in fact refer to the same virtue. A virtue is "a deep and enduring acquired excellence of a person, involving a characteristic motivation to produce a certain desired end and reliable success in bringing about that end" (1996: 137). They are acquired by habituation and "moral work" (125). They are "deep" in the sense that "[w]e think of a person's virtues as closely associated with her very identity" (85). This suggests that they are robust traits; one's identity is not defined by a narrow disposition to act a certain way in a specific circumstance, but by character traits that generally and regularly shape one's behaviour.

Intellectual virtues are reliable dispositions for acquiring true beliefs, just as moral virtues are reliably successful at attaining their ends. Montmarquet has argued against a reliability condition for virtue on the grounds that virtues are not *necessarily* successful. If epistemic vices like partiality, closed-mindedness, intellectual cowardice, and the like were truth-conducive and their corresponding virtues were not, we would not be willing to blame open-minded agents for their virtues and praise vicious ones for their vices. Montmarquet concludes from this that the virtues are constitutive of "free and responsible inquiry" (2000: 140) but only contingently linked to its success. This is a hasty conclusion. It is important that the intellectual virtues actually be truth-conducive, and moreover truth-conducive in relevantly close possible worlds; they need not be necessarily truth-conducive. Possible worlds in which the virtues are not truth-conducive may be dealt with in the same way as standard skeptical scenarios.

Intellectual virtues are marked by the motivation to acquire "cognitive contact with reality" (1996: 167), or true belief, with respect to the proposition in question. This motivation can be very general—a person can be open-minded, for instance, out of a general motivation to understand others. It is only necessary for virtuous belief, however, that one be motivated to believe the truth on the proposition believed. It is not even necessary that one be motivated to believe the truth on the negation of the proposition (2003). This is in response to Sosa's objection that you can know that your parents loved you even if you find the contrary possibility so painful that you would be motivated not to believe it if it were true (Sosa 2001: 51-2). In such a case, you are motivated to believe the truth if your parents love you (which allows you to know that), but are not motivated to believe the truth if they don't. But if you only wanted to believe that your parents love you because all your friends believe their parents love them, and you like following the crowd, you would fail to have knowledge.

As we saw in §II.3, Zagzebski takes knowledge to be an "organic unity" of belief-state and processes of formation and sustenance. She defines knowledge as an act of intellectual virtue, an act that:

(a) arises from the motivation to believe the truth with respect to the proposition known,

(b) leads to a belief that a virtuous person would hold that is formed and sustained in the way that a virtuous person would, and

(c) leads to true (rather than false) belief because of (a) and (b) (2003: 152-3).

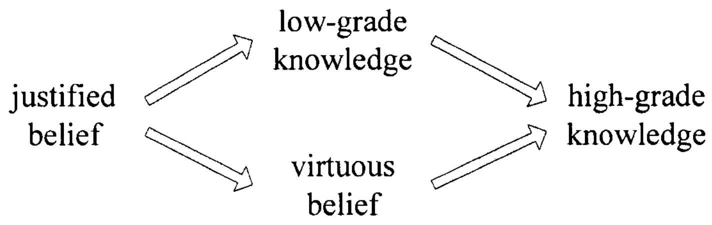
(c) is meant to allow the definition to avoid Gettier cases. In a Gettier case, one believes out of virtue and has a true belief, but the belief is not true because of the virtue. In knowledge, one's having a true belief is due to one's own powers and abilities (1996: 293-9). (This is the genesis of the credit theory of knowledge.) The motivational requirement for knowledge, and the fact that one's belief must be true in part because of that motivation, together rule out knowledge in the cases that are problematic for reliabilism.

Note that actually being intellectual virtuous is not necessary for having knowledge; one must only have the requisite motivation and conform to what a virtuous person would believe. Zagzebski calls the above definition "low-grade" knowledge. "High-grade" knowledge is that which actually involves intellectual virtue, and is a greater epistemic good. This bifurcation in concepts of knowledge is necessary to extend the definition to the low-level virtues. Being faculties and not necessarily acquired by habituation, low-level virtues do not fit Zagzebski's conception. So virtue *per se* cannot be necessary for knowledge. The proportion of knowledge that is high-level depends on the prevalence of high-level virtues and the extent to which they affect the formation of perceptual, mnemonic, and similar beliefs (273-83). Goods such as understanding and wisdom are presumably types of high-grade knowledge.

Zagzebski takes a belief to be justified if it arises from the motivation to believe the truth and conforms to what a virtuous agent might believe (i.e., wouldn't not believe) in the circumstances. This parallels her conception of moral justification. The standards for moral virtue are quite rigorous, and most persons are not morally virtuous agents. But it is implausible that most of us are never morally justified. So to be morally justified, one must only refrain from doing what a morally virtuous person would not do in the circumstances, and act out of the same motivation as a morally virtuous person.

It's plausible to include in the framework a stronger ED associated with belief that actually arises from intellectual virtue (though it may be false or only accidentally true), although Zagzebski doesn't include it in her account. Call this "virtuous belief". That revision yields the system of ED found in figure 5.

Figure 5. Zagzebski's ED framework



3. Non-epistemic virtue theory

As we saw above, the virtues appear to play an important role in reasoning and deliberation. One approach to intellectual virtue examines the role of virtues in facilitating and evaluating reasoning, but leaves open whether the virtues are involved in the analysis of epistemic desiderata like knowledge. This approach is taken by Adam Morton and Christopher Hookway; both are officially cagey about whether the virtues have any role besides regulating deliberation. Thus, their accounts are largely restricted to the high-level virtues.

Non-epistemic theories can be developed on generally reliabilist or responsibilist lines, depending on whether the virtues are primarily individuated by reliable success or by motivational and similar conditions. Morton (as we will see) keeps close to virtue reliabilism. Hookway does not give a systematic account of virtues. But he appears to regard them as essentially Aristotelian, and in describing how they are useful to deliberation he emphasizes their basis in our character rather than their reliability (for instance, in the passage quoted at the beginning of ch. III). For these reasons, he appears to take a generally responsibilist bent.

Morton's (2004) theory is well developed and has some interesting features, so we will take a brief look at it. His chief concern is with how we respond to resource limitations. These (as I noted in §III.2) often make evidentially driven, rigorous solutions to problems intractable, as well as making it impossible to determine in an evidentially driven, rigorous way whether a particular shortcut or heuristic is tractable and likely to be correct in a particular case. The virtues, he thinks, can permit us to solve these problems effectively despite these barriers.

Morton's account of virtues has two components. First, V is a virtue only if

> (a) in some circumstances C, V contributes to the achievement of some ED;
>
> (b) V is less likely to influence cognition in circumstances not in C.

Thus virtues are helpful traits of character that are sensitive to when they are needed; their influence is correlated reasonably well with the conditions under which they are helpful. This way it is not accidental when the virtues do contribute to our cognitive goals.

Virtues also have to have a certain degree of robustness; the circumstances in which they are helpful must be sufficiently broad and varied. This helps distinguish a shortcut that is handy for a single problem or for coping with a narrow range of circumstances from a virtue *per se*. However, for condition (b) to be non-trivial, C's counterpart must be non-empty. This mild robustness requirement and the requirement of sensitivity to reliability might be seen as weak conditions tying virtues to their possessor's character. Morton does not require that virtues be part of character in a stronger sense.

Morton appeals to the virtues to explain how our belief-forming and decision-making capacities can go beyond our capacities to engage in deductive and inductive reasoning from the knowledge we have. Thus he requires that virtues not consist just in propositional knowledge, and not be replicable by the agent's capacities for canonical inference. Virtues may be facilitated by propositional knowledge, but one would expect that they are typically acquired by habituation—by prior experience, practice, emulating others, and so forth. These conditions entail that virtues are not just rules that we can learn and apply blindly. However, when one does not have propositional knowledge as to all and only the cases in which a rule should be applied, the capacity to make use of the rule can be (or derive from) a virtue. So, for instance, knowing how to apply a rule that has a significant number of borderline cases can count as a virtue.

The last two conditions ensure that reasoning involving the virtues is distinct from standard deductive and inductive reasoning. This has the disadvantage of narrowing the scope of Morton's account. If we drop these two conditions, a much wider range of human reasoning involves virtues. For instance, a capacity to use background knowledge to determine how to apply the straight rule would count as a virtue (since it would be a helpful disposition in a wide range of circumstances that is sensitive to its own helpfulness). This would allow for the prospect of making the virtues involved somehow in knowledge and thus unifying the high-level and low-level virtues. But we can still acknowledge that some virtues are interesting because they help us cope with limitations on our capacities to make inferences from our knowledge, and some are dull because they don't.

In this general survey, I have not in any detail described proposals for what makes virtues part of, integrated with, or tied to cognitive character. So let us now turn to a critical examination of the various accounts of that.

V

CHARACTER

In the last chapter, we looked at several different accounts of the virtues. Here, we will examine whether any of these theories provide a suitable account of the relationship between the intellectual virtues and a cognitive agent's own character. We will start by looking at whether the Aristotelian conceptions of virtue preferred by responsibilist theories can do the trick. Then we will turn to the various sorts of virtue reliabilism.

1. Responsibilist character

Aristotelian virtues

Zagzebski's account of how high-level virtues are connected to cognitive character is essentially Aristotelian: they are deep and entrenched excellences closely connected to their possessor's very identity as a person. The frequency of knowledge very clearly outstrips the frequency of entrenched personal excellences of this sort, and thus she does not take intellectual virtue to be a precondition for knowledge.

It will be worth spending a moment on why this is the case. Social psychology has found that a person's behaviour is highly sensitive to the situation that she is in and often inconsistent between situations. Robust character traits have little value in predicting behaviour, while situational elements—even apparently trivial ones—are often excellent predictors. (See Doris 1998 for a summary of the relevant evidence.) On these grounds, it is often argued that virtue theorists are mistaken in basing ethical evaluation on robust character traits that the vast majority of persons do not have (Doris 1998, Harman 1999).

The argument is a bit quick, however. In every study, there are always a few subjects whose behaviour is not predicted by situational factors. These persons may exhibit robust character traits. Social psychology would thus tell us that virtues are perhaps rarer than we think, but that does not mean that they do not exist, and it certainly does not mean that robust virtues are not a moral ideal to aspire to (Sreenivasan 2002).

Much the same situation can be expected to obtain for our cognitive capacities. If behaviour generally does not exhibit robust traits like courage, humility, and so forth, then we cannot expect human thinking to often exhibit the intellectual analogues of these. What we do find is that cognitive capacities often have narrow scope and are highly sensitive to "contextual" factors such as environment and the way that problems are presented (Ceci 1993, 1996b). For instance, one study found that uneducated maids engage in very accurate proportional reasoning when buying goods, but only approximate reasoning when cooking. Construction workers can solve geometrical problems when presented as problems that would arise when doing construction, but not when these are presented as problems that would arise when working in a juice factory or presented in abstract form (see Ceci 1993: 422-7 on both studies).

Such studies do not disprove the existence of robust traits of cognitive character. They certainly suggest that there are intellectual virtues for recognizing when a strategy that is successful in one context can be applied in another. They also show that we can sometimes satisfy our cognitive goals using highly nonrobust, narrow strategies or processes. Robust virtues might be the ideal, but they cannot be prerequisites for any common epistemic desideratum, like knowledge.

I won't try to estimate the extent to which human cognition depends on narrow, context-dependent skills rather than broad traits. What is important is that we cannot rule out the existence of context-dependent skills *a priori*. We cannot even rule out their prevalence. It is important that our epistemology not assume that cognition can be adequately described by broad-ranging faculties like "coherence-seeking reason", rather than by disparate collections of narrow processes. It is also important that we not tie the subject-attributability of epistemic desiderata to generation by broad-ranging or robust cognitive traits. That position runs the risk of inadvertently ruling out many clear cases of knowledge. We are only in a position to take analogues of Aristotelian moral virtues to be involved in relatively rare (but perhaps especially valuable) epistemic desiderata.

Motivations

Zagzebski argues that to have knowledge is to have a true belief because one is motivated to believe the truth and conforms to the belief-state and acquisition process of a virtuous agent. It is not clear in what sense one's belief and belief-formation are supposed to be the same as a virtuous agent's. Presumably, what is intended is something like this: a virtuous agent in the same evidential situation, with approximately the same cognitive capacities (except, of course, for the virtue), would form the same belief using the same type of process. However, this formulation contains several underdefmed concepts with no clear indication as to how they are to be developed. Thus the motivational requirement does the real work in picking out knowledge. The motivation to believe the truth is parallel to a subjective justification requirement like the demand that agents use dispositions they countenance. But since it is also an essential component of intellectual virtue, it bridges the gap between a true (and reliably formed belief) and the believer's character—in this case, his personality and desires.

Clearly, one can have knowledge without being motivated to believe whatever is the truth on the matter. Suppose I want to believe that my daughter is very smart; suppose, moreover, that she is[1] and I have good evidence for this. Clearly, I can know that she is very smart despite my motivations. Zagzebski (2003) argues, however, that I am still motivated to believe the truth if my daughter is very smart. If she weren't, of course, then I wouldn't be motivated to believe *that* truth. But given that my daughter is very smart, my motivation not to believe the truth if she weren't does not affect the formation of my belief and does not prevent me from knowing it. If, however, I didn't really care if my belief that my daughter is very smart is true—if, say, I just wanted the pleasure of believing it, rather than wanting it to be the case—then I would not even be motivated to believe the truth in this weaker sense, and could not know the belief.

However, one can know that p without being motivated to believe the truth if p is true. Suppose Jacob is very strongly motivated not to believe that his wife is cheating on him. For months, he has been ignoring contrary evidence, devising convoluted interpretations of her behaviour, and preventing himself from drawing the logical implications of the evidence in order to keep from having to form that belief. Then, one day, he walks in on her and her lover *in flagrante delicto*. Surely he knows that she is cheating on him once he has seen it with his own eyes. Yet it is not plausible that his motivation not to believe has suddenly gone away. For instance, he might still be inclined to consider alternative explanations that would allow him to continue deluding himself. Of course, if there are no such explanations in relevantly close possible worlds, then Jacob's motivation to believe them could not influence his belief-formation.[2]

Jacob can know despite not wanting to know because our belief-forming processes are to a certain extent autonomous. There are limits to the extent that

[1] Which is, incidentally, true.

[2] In Machiavelli's play "The Mandrake Root", a wealthy old man is tricked into allowing a playboy to seduce his young wife. The couple has had difficulty conceiving. The old man is told that his wife will become pregnant if she drinks mandrake-root tea, but the first man who sleeps with her after she drinks it will die. And so the husband is persuaded to allow another man to precede him.

We may suppose, however, that in no relevantly close possible worlds is Jacob *this* gullible.

our motivations can influence them. Given the pervasiveness of the tendency to believe what one wants to believe, it is almost certainly a good thing that our belief-forming capacities are partially insulated from our motivations. It is important for us to be able to form beliefs that we do not wish to have, and even *know* them despite preferring not to.

Of course, had Jacob been properly motivated, he could have learned the truth much earlier. Proper motivations do clearly play an important role in virtuous belief-formation. This might lead one to think that we can keep the motivational component of high-level virtues, and merely drop the motivational requirement for knowledge. But then it would be entirely unclear in what sense virtues are involved in knowledge. One might be able to build on the vague requirement that knowers form beliefs as the virtuous agent would despite not actually having the virtues, but I do not know how that could be done.

There seems to be an easier way to explain how motivations are involved in virtue. Bad motivations tend to mislead us, by making falsehoods seem attractive, dulling our sensitivity to contrary evidence, or discouraging us from engaging in inquiry and deliberation that would correct our beliefs. The motivation to believe the truth encourages dispassionate evaluation of the evidence as well as energetic inquiry and criticism. So then the motivation to believe the truth is instrumentally valuable for virtuous inquiry, because it helps us achieve our cognitive goals. Motivations to believe contrary to the truth are bad because they tend to lead to unreliable belief-formation.

The link between bad motivations and unreliability is close enough that we usually interpret evidence of the former as an indication of the latter. Suppose, for instance, that you are told that drug D has been found to be safe by a study funding by the company that makes drug D, and that no one has yet attempted to corroborate this claim independently. Most people will not take the study as sufficient reason to believe that drug D is safe because it is too likely that the researchers were motivated to find D safe rather than to discover the truth.

If Aristotelian accounts of virtue sever the putative link between virtue and knowledge, then we are going to have to look elsewhere for an account of how virtues are linked to cognitive character. We thus now turn to virtue reliabilist accounts of character.

2. Reliabilist character

Sosa says that an agent's inner nature is "a *total* relevant epistemic state, including certain stable states of her brain and body" (1991: 285, e.i.o.); elsewhere, he calls it "intrinsic" (141) to her. I will give a few arguments here that should indicate that a rough, underdeveloped conception along these lines won't do any serious epistemological work, and that the problems go deep enough that it is unclear how to develop the notion to avoid them.

I take it that one's "inner nature" must have two characteristics. First, it must be internal to the agent. Let us understand this broadly, and suppose the "internal" to mean either internal to the mind—including one's full set of beliefs and desires, experiences, and so on—or internal to the physical body. Second, it must be part of some subset of the internal that is "natural" in a sense in which, at least, brain tumours are not natural. I'll look first at the possibility that virtues must be part of one's inner nature. That will turn out to be a dead end; then we will turn to the possibility that there is a sense in which virtues derive or arise from one's inner nature.

Realization in inner nature

Let us start with the internal aspect, and understand it as meaning that virtues must be dispositions to believe correctly that are realized in, or supervene on, the internal. That is, let us suppose that virtues must be dispositions of either immaterial minds or agents' bodies. Since immaterial minds are mysterious entities, let us focus on the latter idea.

The problem is that it appears that cognitive processes can be realized in part outside the physical body. Consider, for instance, doing complex symbolic reasoning—long division, say. Most humans do not have the memory capacity to keep track of the results from all earlier stages of the calculation. We use pen and paper as an external memory store to help keep track of all the necessary information.

The idea that cognitive processing can extend outside the mind is generally called the dynamic-embodied (DE) approach (e.g., by Hurley 1998) or wide computationalism (Wilson 2003). The position has been argued for extensively elsewhere,[3] and it would be too long a digression for me to reproduce those arguments here. Here, I will focus on the example I just gave, and argue just for the conclusion that this one type of belief-forming process is realized outside the physical body. That is all we need for my purposes here.

There are two ways of theorizing about pen-and-paper calculations. First, we could suppose it to be a sequence of disconnected cognitive processes. Here is the rough idea. Suppose you are computing 293 + 304 + 57. The first process in this sequence consists of determining that the numbers need to be written in columns like

$$\begin{array}{r} 293 \\ 304 \\ \underline{\ 57} \end{array}$$

and initiating motor actions to do this. The second process consists of visually retrieving the ones digits of each number, summing them in the head to get 14, and then initiating motor actions that yield

[3] See, e.g., Clark (1997), Hurley (1998), and Wilson (2003).

$$\begin{array}{l} \scriptstyle 1 \\ 293 \\ 304 \\ \underline{57} \\ 4 \end{array}$$

This continues for several more steps. The penultimate process finishes with writing the last '6' on the paper, yielding

$$\begin{array}{l} \scriptstyle 1\,1 \\ 293 \\ 304 \\ \underline{57} \\ 654 \end{array}$$

and then using vision to form the belief that the sum of the three numbers is 654.

The second approach is to treat the entire process of calculation as a single computation, rather than a sequence of separate computations. This is necessary for understanding why that sequence, rather than another one, occurred—why, for instance, the agent added 3, 4, and 7 in his head rather than some other numbers, and why that came *before* adding 1, 9, and 5 in his head.

Epistemologists have another reason to treat the entire computation as one process leading to the belief that 293 + 304 + 57 = 654. If we suppose that belief-formation occurs only in the head, then we have to treat the belief as arising from the final physically internal stage in the process—namely, looking at the finishing computation and perceiving that the answer is 654. But the reliability with which the belief is formed is *not* the reliability with which one can read a number written on a piece of paper. It is determined by one's ability to perform the entire calculation. Consider an agent who never misreads numbers and can correctly identify the output of a sum written in that form, but who is almost always wrong when summing numbers himself. If we individuate his belief-forming process as just reading the conclusion off the page, his belief that 293 + 304 + 57 = 654 is reliably formed. This is, of course, wrong, but in saying that it is wrong we have to say that the belief-forming process does not just supervene on his body. Since that calculation involves manipulating external structures, we cannot say that it is realized internally. Nor can we say that it is realized in the agent's nature. While it is difficult to say just what human nature is, it is unlikely that it includes pens and pieces of paper.

Derivation from inner nature

Sosa acknowledges a distinction between fundamental and derived virtues, where the latter are acquired through the use of the former (1991: 278). The ability to do sums with pen and paper is a derived virtue both in the sense of being acquired through use of more basic processes, and in the sense of involving more basic processes—most obviously, the capacities to add small numbers inside the head, to recall the correct sequence for adding large sums, and the capacity to read numbers off the paper. So we could suppose that virtues don't have to be realized in one's inner nature, but they do have to be *derived from* one's inner nature.

The most obvious sense in which a faculty is derived from another (or from one's nature) is historical; one is derived from the other by having been acquired or developed through the other's operations. Goldman, for instance, proposes that a knowledge-generating process must but be acquired in a reliable manner—i.e., through processes of acquisition that reliably generate reliable belief-forming processes (1986: 51-3).

There are other ways that virtues could be historically derived from inner nature. On proper functionalism, a true belief is known iff it arises from a faculty that is well-designed and functioning properly in an environment in which it was designed to function (see Plantinga 1993a, Millikan 1993). Our faculties could have been designed by either God or natural selection. If we were designed to acquire a particular faculty—say, by being "natural-born cyborgs" (Clark 2003), designed to incorporate environmental props into our thinking—then that faculty could be said to be derived from our inner nature even if it includes objects that are neither internal nor natural.

The trouble with these approaches is that the history of a process is less important to its evaluative status than its present condition (a point that Sosa has himself emphasized). Consider Swampman, a duplicate of a fully formed, well-educated human who was created (by incredible luck) by a lightning strike in a swamp. Swampman's processes will fail any historical condition on virtues, but it is implausible that this means he could never have knowledge. There are also less strange examples. Greco describes a case (from Oliver Sacks) in which an illness had the side-effect of producing unusually reliable and detailed childhood memories, allowing its victim to produce exceptionally detailed and accurate paintings of his hometown as it was when he was a child. He observes that

> the man was considered to be the foremost expert on the layout and appearance of that town—though he had not visited there in decades. In other words, there was consensus that the abnormality gave rise to knowledge (2003a: 473).

For these reasons, Sosa argues that a virtue must be "a self-sustaining and firm part of one's intellectual character" (Sosa 2001: 57). As long as it has the right properties *now*, it could have originally come to be in all sorts of ways.

So if virtues are derived from inner nature, this can only be in a non-historical sense. The capacity to do pen-and-paper calculations *is* derived from capacities to remember and follow procedures, read numbers, and do simple mental arithmetic in the sense that those capacities are involved in performing the computation. Without those capacities, it would be impossible to perform the calculation. Being internal, these might be supposed to be aspects of inner nature from which the partially external overall process is derived.

However, *any* belief-forming capacity is partially derived from one's inner nature or one's fundamental virtues. A disposition to accept blindly anything the Dalai Lama says depends on capacities for understanding language. A reliable brain tumour doesn't just cause a belief by itself; it needs, at a minimum, the conceptual capacities to believe in one's impending death. These other capacities are quite plausibly part of (or derived from) one's inner nature.

This doesn't conclusively establish that there is no way of rigging up a sense of "derived" on which intuitive candidates for virtue are derived from inner nature and intuitive nonvirtues aren't. Perhaps we could say that the agent's inner nature must be the most salient element in the production of the belief. This threatens, though, to be a shallow analysis: it might tell us how we distinguish clear cases of virtue from clear cases of non-virtues without telling us what the real difference between the two categories is.

Cognitive integration

Greco's notion of "cognitive integration" shows promise of getting around the problems we've seen with Aristotelian virtues and inner nature. It does not require that one be motivated to believe the truth. However, it is plausible that dispositions driven by bad motivations would have a tendency to fail to be integrated. Belief-formation is aimed at acquiring significant truths. Inasmuch as a particular disposition is motivated by some other goal, it will be out of step with the rest of the agent's cognition. It's likely that in many cases this would lead to a lack of sufficient integration. Greco does think that integration is in part a function of the range of beliefs generated by a disposition. This could perhaps be problematic for the reasons we saw in §1 above. But a well-developed notion of integration should be able to avoid too much reliance on robust character traits.

Integration does not require internality; a cognitive process that involves manipulating external structures can be just as well-integrated as one that is wholly realized in the brain. Integration is also a matter of the present state of a cognitive system, not its history or intended design. And a full account of integration promises not only to discriminate virtues from non-virtues, but to explain the differences between them. Nevertheless, this is all very speculative; without a proper account of cognitive integration, we don't know if the notion will keep all these promises.

In the next chapter, I'll lay the groundwork for an account of cognitive integration. I will do so by looking at how a bundle of disconnected processes can be turned into a well-integrated agent able to achieve her cognitive goals. This will turn out to be a matter of being able to regulate one's own cognition successfully. In §VII.1, I'll argue that this capacity for regulation is the basis of intellectual virtue.

One note on how we'll conceptualize this. The notion of a disposition to form beliefs is too coarse to make sense of cognitive integration. We will have to be able to talk about the internal structure of dispositions. Thus my focus will not be on dispositions to believe, but on the cognitive processes that underlie those dispositions. The advantage of this is that it will allow us to make use of empirical psychology, and in particular, the study of human metacognition.

VI

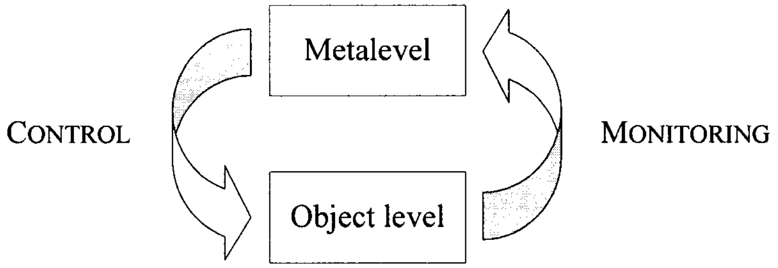
COGNITIVE MANAGEMENT

In this chapter and the next, I will try to answer the question with which we ended the last chapter: how are processes integrated into cognitive character? Very roughly, I am going to propose that this is a matter of having *metacognitive control* over one's processes. A virtue is a capacity to control one's processes so that they allow one to form beliefs reliably and efficiently, so that they allow one to reach one's cognitive goals. The argument for this account of virtue will proceed as follows. In this chapter, I will try to establish that successful cognizers with processes like ours need to control their processes in order to have reliably formed beliefs. Metacognitive control is what makes a bundle of disconnected processes into a system for acquiring reliable beliefs. It thus seems to be exactly what we need to distinguish virtuous, well-integrated processes from the rest. My discussion of metacognition will of necessity be programmatic and abstract, but it should be sufficient for our purposes. In the next chapter, I will develop the resulting account of virtue and give some illustrations of its usefulness for epistemology.

1. Basic principles of cognitive management

Metacognition, as I will use the term, is the monitoring and control of object-level cognitive processes; not just thinking about thinking, but the regulation and management of thinking. Like most phenomena worth examining, it is easier to recognize than to define. Nelson and Narens (1990) propose three principles that characterize metacognition and that seem to be broadly accepted in the field. First, cognitive processes are split into an object level and a metalevel. Second, the metalevel is a model of the object level (and not the other way round). Third, the two levels are connected by relations of monitoring and control, as illustrated in figure 6. In monitoring, information flows from the object level to the metalevel and informs the latter's activity. In control, the metalevel exerts causal relations on the object level that regulate its behaviour by initiating, sustaining, or terminating activity at the object level. This can include preventing object-level activity from eventuating in actions or beliefs, or permitting it to do so.[1]

Figure 6. The general structure of metacognition



The first principle says that the object level and metalevel cannot be identical. It is important to note that the distinction between levels can occur on a case-by-case basis. It does not commit us to thinking that all metacognitive functions are performed by specialized "higher" processes, or that there is a stable category of object-level processes. For instance, a metaprocess evaluating the chances of success of different problem-solving strategies might make use of the

[1] "Process", here, refers to something narrower than a belief-forming process. Most belief-formation involves both object level and metalevel processes. In ch. VIII we'll look at how processes are individuated.

brain's usual centres for making inductive inferences.[2] But these processes themselves would be object-level with respect to metaprocesses that prevent statistical blunders or the projection of unprojectible predicates.

The second principle says that the metalevel functions as a model of the object level; i.e., that there is a mapping from events at the object level to responses at the metalevel.[3] Consider the behaviour of the governor on a steam engine. Assuming that the governor is optimal—responds immediately to changes in the engine's speed, etc.—there will be a mapping from the speed of the engine to the rate of intake the governor allows. Thus the behaviour of the governor functions as a model of the engine's speed. An imperfect governor—one that is insensitive to certain changes in speed, or slow to respond—imperfectly models the engine's speed. We can expect, of course, that human metacognition is generally imperfect. But we should nonetheless expect metalevel events to model object level events approximately.

It is important not to confuse the metalevel's *being* a model of the object level with its *constructing* a model thereof. A perfect governor is a model of its engine's speed, but it nowhere represents the engine's behaviour. Constructing models is difficult and resource-intensive. When information can be quickly retrieved from the external world, it is typically more efficient to access it only when needed for processing. Andy Clark dubs this the *007 Principle*: "know only as much as you need to know to get the job done" (see 1997: 46); more precisely, represent only what you need to represent to get the job done.

There is evidence that the human mind is constructed in accordance with the 007 Principle. For instance, we do not construct very detailed models of the visible environment, as the phenomenon of "change blindness" shows. (See O'Regan & Noë 2001: 954 for discussion.) Low-level visual processes monitor

[2] This is one possible explanation of the correlation between success in strategy selection and aptitude at inductive reasoning (see Schunn & Reder 1998).

[3] The principle is based on Conant & Ashby's (1970) theorem. Let R be the simplest optimal regulator of a system S. Let $\sigma(i)$ be S's response to input i and $\rho(i)$ be R's response to input i. Then there is a mapping h: $S \rightarrow R$ such that $\forall i$, $\rho(i) = h[\sigma(i)]$.

the environment for signs of change (such as unexpected changes in retinal stimulation). These processes can be neutralized by, for instance, having a visual scene flicker when a change is made to it, or by making a change during a saccade, or by having the subject attend to something else. Then, drastic changes can be made in the visual field without subjects noticing. Brooks (2002: 82-3) describes a study conducted by Ballard and Hayhoe that illustrates this beautifully. They presented a pattern of coloured blocks on one side of a screen and had subjects reproduce it on the other by picking up and arranging blocks with a mouse. When the subjects was picking up a block and not looking at the original pattern, the investigators would change the colour of blocks in the original pattern. Subjects completely failed to notice that these changes had occurred.

If we produced detailed internal models of the visual scene, it would be easy to compare incoming information about the present block pattern with the present model and thus determine that the two are different. That we cannot necessarily do so indicates that we do not produce detailed models of information that we expect our visual system can readily retrieve directly from the environment if necessary.

This does not mean that the human mind does without representations altogether, of course. Some problems are "representation-hungry", requiring representations for successful computation (Clark & Toribio 1994). Some such problems involve absent, counterfactual, or physically disparate classes of objects (Clark & Toribio 1994); others, like planning rapid motor actions, must generate outputs more quickly than the environment can supply feedback (Grush 2003). Nonetheless, when cognitive processes can be directly coupled with the environment, there is no reason to expect extensive modeling to occur. Since metacognitive processes are often directly coupled to their objects, we should expect that at least some would be able to forego the construction of explicit models.

2. Types of metacognitive problems

For epistemological purposes, what is most important is the type of problems that metacognitive processes are needed to solve. Roughly, these are problems involving the coordination and integration of different processes or strategies. They thus cannot be solved (though they can be facilitated) by independent object-level processes acting alone. In this section, I will examine three functions for which successful epistemic agents need metacognitive processes: *conflict resolution*, *selective application*, and *resource management*. With these three, we can establish the importance of metacognition in the cognitive system, which I will build on in the remainder of this thesis. Metacognitive control may be important in other ways as well; to say the least, that would not weaken my case.

Conflict resolution

Let me start with some facts about belief. Belief-states have extensive connections with other mental states. They are involved in the causation of actions—*ceteris paribus*, if you believe that p, you will act in ways that will satisfy your occurrent desires if p is true. They are involved in the generation and extinction of desires—*ceteris paribus*, if you desire A and believe that B will bring about A without undesirable side effects, you will desire B; whereas if you desire A and believe that A will bring about side effects that outweigh getting A, you may cease to desire A. Beliefs are also involved in reasoning, since, *ceteris* very *paribus*, we reason in ways that would be truth-conducive if our beliefs are true. Beliefs are involved in assertion—*ceteris paribus*, if you believe that p and wish to be honest you will be disposed to assert that p. And so on and so forth.

This picture is heavily qualified by considerations of resource limitations, self-knowledge, etc. Though these connections can be weak or suppressed and depend on the subject's other psychological states, one general point emerges—

believing is a state involving, and realized in, large parts of a cognitive system. Properly speaking, it is an attitude that an agent or an entire cognitive system takes. Believing is not just a matter of having information stored somewhere in the system, say, in long-term memory or in a perceptual buffer. The information must have the requisite connections to other parts of the system in order to be a belief. These connections are difficult to spell out for the general case, since they depend on the rest of the subject's beliefs and desires. Nonetheless, without such connections, there is nothing that makes the content of that information *believed*, as opposed to being subject to another psychological attitude—being, say, entertained, or seeming to be true without really being believed, etc.

While I do not intend to give a full analysis of belief here, I should say something about how to distinguish belief-states from other attitudes. This is particularly pressing since, with all the different connections that beliefs have, there is a wide variety of borderline cases of belief that have some of these connections but lack others. (See Morton 2003: ch. 3 for discussion.) Delusions are an interesting example. It appears that they can arise inferentially, and they are of course tied to assertion, but they often do not cause actions in the ways that ordinary beliefs would. For instance, victims of Capgras delusions, who believe that a loved one has been replaced with a look-alike imposter, generally do not report the imposters to the police. (See Davies & Coltheart 2000.) For epistemological purposes, it is natural to identify belief-states as states that can be subject to epistemic appraisal. In particular, a belief is a state that one can legitimately say is known or not known. This standard is like Williamson's (2000: 41-8) analysis of belief as attempted knowledge, but is a purely operational criterion. It helps identify the states that are relevant for epistemology, and here is not the place to decide if it means any more than that.

The one wrinkle with this method is that sometimes we do use "know" in a sense that does not imply belief. One might say "I knew it was the house on the left" when one considered but rejected that proposition. Such cases can be ruled out, however, by noting that it is also true to say "I should have known". Only one of "x knows" or "x should have known" can be true if "knows" is used in its literal sense.

The criterion excludes, e.g., perceptual experiences (which Goldman 1986: 185 calls a type of belief) or information retrieved from long-term memory, when we do not consider these veridical. Merely seeing the stick bend in water or seeming to remember paying the phone bill cannot meaningfully be said to count as knowledge or not; only after the subject accepts the experience or the retrieval as veridical or not is the resulting state a candidate. I also take delusions not to be beliefs by this standard. Although they are surely belief-like, we do not normally apply epistemic appraisals to them. They are *beyond* unjustified, not known, ill-formed, and so forth.[4]

This conception of belief is, however, broader than "acceptance" (Lehrer 1990) and related conceptions. It includes beliefs that are qualified in various ways and beliefs that the agent is unaware of holding or would even vociferously deny. A person who treats minorities in discriminatory ways while honestly averring a fervent opposition to racism can be criticized specifically for epistemic laxity. (On the other hand, a person who disavows racism but is uncomfortable around members of certain groups is not epistemically culpable.) Interestingly, it also includes dispositional states that the subject cannot at the moment retrieve—it is perfectly legitimate to say "she knows that p" or "she knows she shouldn't do that" when in that specific circumstance p cannot be retrieved or temptation has overridden moral knowledge. This conception includes faith and other states that can obtain without adequate epistemic reasons (*contra* Adler's 2002 conception of belief). Consciously appraising that one does not have sufficient reason for thinking a belief true is uncomfortable and at least in the philosophically astute

[4] It might be objected that this element of common practice is merely due to a lack of full appreciation of the similarities between delusions and ordinary belief-states. Only simplicity of exposition turns on denying delusions the status of beliefs. Since no delusions are knowledge, we would have to posit a minimal rationality requirement on beliefs to ensure that they were not delusional. The discussion would be otherwise unchanged.

precludes full acceptance. Nonetheless, it is quite legitimate to evaluate states without reasons as known or not known, and thus by this standard they constitute beliefs.

All the states that are here counted as beliefs are interconnected with other mental states and events in ways that place constraints on the extent and type of conflicts that can be found among an agent's beliefs. To exist in the same subject, inconsistent beliefs have to be insulated from each other in various ways. Like Frege, one might fail to draw out the implications of one's beliefs far enough to discover the inconsistency. One might even unconsciously avoid drawing implications from certain beliefs, in case that would show them inconsistent. Or inconsistent beliefs might become occurrent at different times or with different prompts. I might believe that I will be home by six and also believe that I will be much later than that, provided that the latter occurs when I think of how much work I have to do and the former occurs when I talk to my girlfriend. There are more complicated strategies for maintaining inconsistent beliefs. A person might be able to assert that not-p honestly but act as if she believed that p by inventing rationalizations for her actions that indicate why they stem from beliefs other than that p. Inconsistent beliefs that are appropriately qualified can even occur at the same time. Consider a physicist who believes all three of classical and quantum mechanics and general relativity. She can know that all three are inconsistent, yet still hold qualified belief in all three because she only applies one theory to any given problem, and the set of problems to which each applies is sufficiently disjoint to avoid contradictions in practice.

Despite all this, *some* inconsistencies among beliefs are not possible. This is particularly true for occurrent beliefs, where the opportunities for successful insulation are much rarer. The problem is that belief-states are tied in myriad ways to action, inference, assertion, and so on; incompatible states can inhibit each other's connections to other psychological states so that neither can qualify as a belief. Suppose, for instance, that the belief that p plus the rest of the agent's beliefs would lead her to do A, and the belief that not-p plus the rest of her beliefs would lead her not to do A. (Suppose, for instance, she wishes to honestly assert whether she believes p.) If both belief-states occurred at once, she would presumably have to both do A and not do A; outside of Zen koans, that simply isn't possible. She might be inclined to believe in incompatible directions, but this is not the same as actually having conflicting beliefs. If attitudes to contradictory propositions sever enough such connections, then neither can be properly said to be a belief-state.[5]

Cognitive processes generating belief-states conflict with each other when they provide the sort of incompatible data that cannot all be believed at once. When this happens, a certain amount of central control is necessary to adjudicate between the conflicting processes and actually generate a belief. Thus it is a task for which metacognitive processes are necessary; without them, one cannot take meaningful doxastic attitudes in cases of conflict. The point may be made clearer with a concrete example. Take BonJour's Norman (see §II.2), who is a model of the supposed gulf between reliability and responsibility. Norman has reliable clairvoyant powers that sometimes produce beliefs in the face of the evidence he has. Suppose Norman's clairvoyance does what it does to produce the belief that the President is in New York City. Suppose further that at that moment Norman is in the White House, looking at a man who looks just like, and has just been introduced as, the President. If Norman's clairvoyance can by itself produce the belief that the President is in New York, then his perceptual and inferential faculties should by themselves be able to produce the belief that the President is *not* in New York. But Norman cannot believe both of these simultaneously. If he thinks he is meeting the President he will say, "Pleased to meet you, Mr.

[5] Delusional persons are sometimes described as having inconsistent occurrent beliefs—for instance, believing at the same time that the same person is both dead and in a room down the hall. Many such manifestations of beliefs are not inconsistent in the way I describe above. It is perfectly possible to be simultaneously disposed to honestly answer "Yes" when asked whether p and when asked whether not-p. It's just not possible to be disposed to answer both "Yes" and "No" when asked whether p.

President," and he wouldn't do so otherwise. Since his clairvoyance and his other faculties are at an impasse, neither one can force a course of action characteristic of belief in either of the inconsistent propositions. What he believes—even whether he adopts a belief-state at all—will be a function not just of a clairvoyance-module or of perception- and inference-modules, but of the way his cognitive system as a whole resolves the conflict.

It's worth noting here that Sosa takes conflict-resolution capacities as indicative of reflective knowledge, which "manifests not just modular deliverances blindly accepted, but also the assignment of proper weights to conflicting deliverances, and the balance struck between them" (2004: 291). The arguments here should emphasize how, except for the emphasis on conscious reflection, reflective knowledge is not far from what we would normally think of as creditworthy cognitive activity. Adjudicating between conflicting deliverances is not just a particularly valuable capacity, but is necessary for wide-ranging reliable belief-formation at all.

Information conflicts are a persistent feature of human cognition. Suppose I am walking down the street in Edmonton and see what appears to be my mother drive by in her car. My background knowledge then reminds me that my mother lives 2300 miles away. In order to form a belief, I must somehow resolve this informational conflict. Or suppose the phone bill appears to say that I still haven't paid last month's bill, when I seem to remember having paid it. To form a belief, I must determine whether my reading, my memory, or the phone company is deceiving me. In either case, in order to form beliefs *reliably* I must have capacities that reliably indicate when a process that is in conflict with another should not be trusted, and what beliefs may appropriately be formed in the situation.

Selective application

A second need for central control of belief-forming processes in agents like us arises from the fact that the reliability of cognitive processes with which we are familiar varies tremendously across different environments. Vision is not reliable in a funhouse or at twilight in a landscape littered with barn façades; hearing is not reliable under water or on the moon. Conduits also vary in reliability depending on the content of the information: your sense of smell will not inform you of the presence of carbon monoxide, and medical science was greatly advanced when it was realized that something can be dirty even though it looks spotlessly clean. To acquire beliefs reliably, agents must have the capacity to *selectively apply* their cognitive processes: to be able, most of the time, to use cognitive processes to form beliefs only in environments in which and for contents for which the process yields true beliefs (Lepock 2006).

As with conflict resolution, the need for selective application is a persistent feature of our cognitive lives. We have already seen a few perceptual examples. One important feature that we use in determining whether perceptual processes can be trusted is the vividness and detail of the data they provide. But there is more to perceptual monitoring than that, since we base beliefs on the absence of data as much as on its presence. Ordinarily, lack of auditory stimulation tells us that there are no large moving objects in the vicinity, but not when one is under water or wearing earplugs.

Many faculties take doxastic inputs, and are not reliable unless their inputs were reliably formed. Memory is an obvious example; selective application is necessary to ensure that retrieved traces originally derived from reliable sources. This is made a more pressing problem by our abilities to reconstruct memories through inference, embellishment, and integration with other beliefs. This dramatically increases the power of human memory, but means that its reliability depends on our ability to distinguish reconstructed memories from fabrications (Mitchell & Johnson 2000).

The same goes for deduction. The fact that one's beliefs entail a conclusion does not necessarily mean one ought to believe it; sometimes this is a sign that one or more premises should be abandoned. Reliable deductive capacities require being able to tell the difference between an interesting consequence and a reductio.

Inductive reasoning especially requires selective application because many, perhaps most, ways of applying the straight rule are not truth-conducive. Gruesome predicates are obvious cases, but the problem arises for less strange predicates as well. Take a woman who is nine months pregnant with her first child. One cannot reason that since every day she has been alive so far, she has not had a baby, she won't have a baby tomorrow. Successful inductive practice requires extensive background and great care in determining what predicates are projectible when; i.e., successful inductive practice requires selective application.

These examples all involve the capacity to refrain from using processes when they would not be truth-conducive. Another aspect of the capacity to selectively apply processes is to be able to initiate them when they are likely to form (desired) true beliefs. When apportioning study time to material, one crucial consideration is to spend enough time reviewing difficult material to be able to recall it accurately; this requires exposing oneself to the material to the extent necessary for understanding and retention (see Nelson & Narens 1990). (The other part of the problem is apportioning limited study time so that one can cover all the necessary material, a type of problem discussed immediately below.) This positive aspect of selective application is less relevant to reliability than to power; it is a matter of extending the range in which one can make fruitful use of one's processes, rather than avoiding using them when it is not fruitful to do so.

Resource management

Real agents have limited cognitive resources—e.g., working memory and attentional capacities—and limited time in which to form significant beliefs. Thus it is important for them to have the ability to apply their processes *efficiently*: to be able to select strategies and initiate processes that will lead to true beliefs while using minimal cognitive and temporal resources. This isn't just a matter of trying to maximize power while minimizing effort. Rather, the problem is that without careful cognitive management, we cannot form the beliefs we need at all. Creatures who distinguish red lights from green by explicit logical deduction from sense-data have a praiseworthy tendency to die without reproducing.

Resource management overlaps with the two functions already described. Effective management of resources requires being able to determine not just whether a faculty can be trusted, but whether it will produce a trustworthy answer at all, in order to avoid the pitfall of initiating processes that will not terminate with a suitable output. Suppose, for instance, that you are trying to remember whether one heard a particular sentence on the news. There are two available strategies for solving the problem. You can determine whether the sentence's content is plausible given what one remembers of the content of the news story, which is a moderately reliable strategy. Alternatively, you can try to remember as much of the story as possible and see if you retrieve the sentence in question. This strategy is quite reliable immediately after watching the news, and drops as a function of elapsed time afterward. Effective resource management involves determining which strategy is more likely to yield a correct answer, in order to avoid wasting time estimating the plausibility of something that could be retrieved, or trying to retrieve something no longer available.[6] But, of course, determining this is also part of selectively applying one's processes.

When processes operating with insufficient resources do yield an output, the result is often likely to be erroneous. There is a variety of different techniques for multiplying numbers in one's head. Multiplying any given two numbers may require choosing a method that, for those particular numbers, will not tend to overwhelm one's resources and lead to miscalculations. Thus in these cases careful management is important for acquiring a reliable answer.

[6] See Cary & Reder 2002 for discussion of the experimental paradigm.

The difficulty of performing each of these three functions will depend on the nature of the processes involved. In humans, resource management is easy for much of visual processing, which is fast and does not compete with other processes for resources; and for recognizing whether something has been experienced before, which is nearly instantaneous and seemingly effortless. At the other extreme, consciously weighing the evidence for a proposition is often best left to the fireside.

More importantly, while ordinary human processes can readily lead to false beliefs if trusted in the wrong circumstances, there are possible processes that would be trivially easy to selectively apply. For now, let us assume that we are dealing with agents with processes that, like ours, must be selectively applied to yield reliable beliefs. The alternative possibility deserves special discussion, which I will defer to the next chapter.

3. The multiple realizability of metacognition

Now that we have seen some examples of metacognitive problems and why solving them is important for being a creditworthy epistemic agent, let us look at just how general the notion of metacognition is. Metaprocesses should coordinate and integrate underlying processes. There are few restrictions on the methods they can use to do this. Metacognitive computations can be realized by procedures of many different sorts.

First, there is no necessary link between managing one's own cognitive behaviour and doing so consciously. An unconscious process could, for instance, monitor the success rates of different problem-solving strategies and initiate the one with the highest chance of success just as easily as a conscious one. In humans, in fact, unconscious metacognitive processes have distinct advantages, since they can evade the extreme slowness and limited processing capacity of conscious deliberation.

Consciously accessible metacognitive processes are easier to study because subjects' introspection can be a fruitful source of data.[7] They thus occupy a central place in the literature. Furthermore, qualia seem to play important metacognitive roles. The qualitative aspects of experience tell us a great deal about the source of our beliefs and how we could gain more information about their objects. A visual image of a pen on a table indicates not just that there is a pen on a table, but that this belief was acquired through vision and that operations leading to more or better visual information can be used to learn more about the pen. Similarly, the qualitative experience of memorial retrieval, the "feeling of knowing", plays a very important role in determining whether to accept the retrieved data (see Koriat 1994).[8]

Some authors do use the term "metacognition" to refer specifically to conscious regulation[9] (e.g., Darling et al. 1998, Koriat 1994), but this is just a narrower usage; it appears that strategy control can be unconscious. In a number of experimental situations where subjects have a choice between different problem-solving strategies, they tend to choose the strategies with the highest base rates of success. But most subjects are unaware of the base rates of success of different strategies, and many are even unaware of the fact that they were using different strategies on different trials (Cary & Reder 2002). Thus, although consciousness may play an important metacognitive role, nonconscious processes appear to perform similar functions.

Talk of metacognition and metacognitive processes should also not be

[7] Which is not to say that introspective reports must always be taken as accurate. They provide data to be explained; often, but not always, the simplest explanation takes them to be correct. See Nelson & Narens 1990.

[8] Note that this assumes that qualia are not epiphenomenal. If subjective experience can't have causal influence on object-level processes, we should wonder whether consciousness plays any metacognitive role at all. The experience of a pen on a table might be a sign that appropriate systems have been informed about where the belief about the pen comes from and what can be done to get more information, but any actual management would have to be occurring beneath consciousness.

[9] In the developmental literature (e.g., in Paris 2002), the term is even sometimes used just to mean "thinking about thinking", including, e.g., thinking about what others think of you.

taken to involve any commitment to particular structures to be found in the brains or equivalent thereof of cognitive agents. A metacognitive process, for my purposes, is just any process that fits Nelson and Narens' three principles (especially if it solves one or more of the problems I described in §2 above). There is no reason to think that any one metaprocess performs a wide range of metacognitive functions. The requisite functions might easily be accomplished piecemeal, as long as each aspect is captured by some process or other. Most certainly, we should not think that some supreme faculty arbitrates between processes like a court at law. Rather, we should play it safe and think of metacognition as a disparate bundle of different processes performing different tasks that provide a potentially very uneven sort of central control. Good cognitive management need not be authoritarian.

For instance, it is possible to manage conflict resolution by having all belief-forming processes feed their results into a single network of propositional representations; this could automatically balance out the various inputs by reaching a stable and coherent pattern of activation that represents the final belief-state. Although there is no "higher" metaprocess actively regulating the object level, this sort of passive regulation of belief-formation by coherence is nonetheless regulation. In principle, there is no reason why such an architecture could not resolve conflicts in a suitably reliable manner, in which case it would surely count as appropriately weighting the deliverances of different processes under different conditions.

In fact, although we have spoken of metacognitive processes, metacognitive control need not be internally instantiated at all. External behaviours can serve the same function. For instance, one cannot will one's visual processes to yield a sharper image. But squinting, manipulating the amount of light, or moving one's body can allow one to improve the visual image and thus form beliefs that would not have otherwise been possible. Similarly, jogging one's memory is a matter of trying to initiate a retrieval process that is not under direct control of the will. Such behaviours serve metacognitive functions; they can contribute to successful cognitive management despite not being, strictly speaking, cognitive processes.

For another example, suppose S tends to find himself seeming to remember as being true statements that he never believed, but to which he was repeatedly exposed. (Humans are generally like this—see Hasher 1977.) S can prevent this limitation from giving him false beliefs by taking various measures to limit his exposure to sources making unreliable statements, by choosing his associates well, his choices in reading and television, and so forth. He should then count as having exerted indirect metacognitive control over his faculties, which is not necessarily any less effective by virtue of having taken a causal detour through the environment.[10]

Let me conclude this chapter by mentioning an important feature of metacognitive monitoring: all that matters is whether it permits the system to manage itself successfully. I have already noted that metaprocesses need not build detailed models of their objects. In fact, it is not even strictly necessary to directly monitor the trustworthiness of processes; all that matters is that whatever is monitored or represented allows the system to control the involvement of that process in reliable belief-formation. For instance, the more steps an arithmetical calculation or a string of deductions has, the more likely it is that errors will intrude. A significant part of the task of controlling such processes could be accomplished by heuristics that only make use of the length (or predicted length) of computations, without actually represent the trustworthiness of the outcome.

Since metacognitive monitoring is judged only by its contribution to control capacities, there are no *epistemic* restrictions on its etiology, content, form. In particular, agents need not have well formed beliefs, or beliefs at all, about their processes. Reflection on one's processes can significantly help the

[10] However, there are temporal restrictions on what counts as effective control. Suppose Oedipus's clairvoyance leads him to form beliefs about where his mother is that, in due time, he will discover are almost all false. Prior to this discovery, he does not have proper control of that faculty.

management of one's processes, but is not strictly necessary. This will turn out to be quite important in avoiding regress arguments and the implausibility of perspectival accounts of knowledge, as we will see in the next chapter.

VII

METACOGNITION AND EPISTEMIC VIRTUE

In the last chapter, I introduced the basic principles of human metacognition. I will now apply these observations to the theory of intellectual virtue. I will argue that intellectual virtues are capacities to control cognitive processes so that they facilitate the attainment of one's cognitive goals.

As we've seen, intellectual virtues are ostensibly involved in a host of different kinds of epistemic evaluation. Most importantly, there is the *prima facie* distinction between low-level virtues, which are involved in knowledge, and high-level virtues, which seem to be involved in different sorts of appraisals of agents. To simplify the argument in this chapter, I will concentrate on the virtues as they are involved in knowledge. I'll call these "epistemic" virtues. This will give us a basic conception of intellectual virtue that I will extend in ch. IX to cover the gamut of putative virtues.

Greco takes a virtue to be a belief-forming disposition that is integrated properly with one's cognitive character. On that way of thinking, it would seem that the question I should address here is how virtues are cognitively integrated. The trouble with treating virtues as belief-forming processes is that it seems to rule out any possibility of uniting the high-level and low-level virtues. Open-mindedness, intellectual courage, and the like are not dispositions to form beliefs, though they are (speaking loosely) dispositions to form beliefs or engage in inquiry in certain ways. One forms beliefs by believing the testimony of others, rather than by being open-minded. The open-mindedness is a disposition to treat the testimony of others in certain characteristic ways; it shapes one's secondhand belief-formation though it doesn't produce the beliefs itself.

This way of describing the high-level virtues suggests a possible route of reconciliation between the two levels. We can say that virtues are not belief-forming processes, but capacities to use processes well. Memory is a faculty; the associated virtue is the capacity to use memory well. So we should identify virtues not with cognitively integrated processes, but with whatever capacities are responsible for the integration.

The first part of this chapter will lay out the basic account of epistemic virtue. Then, I will examine the subjective status of virtuously formed beliefs, and argue that they are internally well formed in a sense that is sufficient for knowledge. Finally, we will look at some implications this account of virtue has for knowledge about one's own cognitive faculties.

1. Epistemic virtue

In ch. V, I concluded that to make sense of the notion of intellectual virtue, we have to make sense of how a process can be integrated with an agent's cognitive character. Now a process that can be monitored and controlled for reliable and efficient use is certainly one that is integrated into its possessor's cognitive character. Furthermore, such integration is necessary for having a cognitive character at all, if one starts with processes in need of the metacognitive control described in the last section. A failure to perform any of these tasks places severe limitations on the agent's ability to achieve the goals of her cognitive endeavours. Thus, for processes like most of ours, it seems that integration into cognitive character can be accomplished by metacognitive capacities. That makes those capacities excellent candidates for intellectual virtues.

Of course, it is possible for there to be well-integrated, successful agents whose integration is not due to their control over their faculties. But our model for intellectual virtue is the way we ought to form beliefs, not the way other possible agents do. Clairvoyants, infallible brain tumours, and other remote possibilities become relevant *after* our epistemic standards have been set, when we try to articulate those standards. Thus we should begin with how integration works in the ordinary human case. Then we can determine if the resulting account explains our intuitions about the strange cases.

Given that metacognitive control capacities are necessary for cognitive integration—for using our processes to achieve our epistemic goals—it makes sense to take them to be these capacities to use processes well. Suppose, then, that an epistemic virtue is a capacity to control one's cognitive processes in a way that allows one to form true beliefs and avoid forming false ones. As we saw in the last chapter, metacognitive control is important for attaining power and efficiency as well, but the achievement of these goals is not necessary for knowledge; thus we will ignore that aspect of control for now.

To put the idea more precisely:

> S has an epistemic virtue iff she has a stable capacity to exert control over her processes in a way that allows her to form true beliefs and avoid forming false ones.

We can then also say that:

> S believes B out of epistemic virtue iff:
> (a) B is reliably formed, and
> (b) B's being reliably formed is in part due to S's having some virtue that controls the processes involved in generating B.

In considering this definition, it is important to remember that one can control one's processes even when their operations are unaffected. Consider a plausible visual belief formed from clear, crisp data. Such a belief might be formed automatically without any metacognitive input, since it is formed in circumstances in which vision is highly trustworthy and does not conflict with any other processes. The processes can still be under control if it is the case that were the belief not plausible, or not formed in auspicious circumstances, the agent would not have automatically formed the belief. Control can be manifested just as much in what one could have done, but didn't have to, as in what one did do. Even automatic belief-forming processes can be under effective metacognitive control, provided that they are more or less only automatic in cases where this does not lead to unreliable belief-formation. (We cannot, of course, expect metaprocesses to entirely prevent unreliable belief-formation.) Thus, B can arise from intellectual virtue even if metaprocesses had no causal influence on its formation.

Easy management

We can see how beings like us, with a disparate bundle of differentially trustworthy processes and limited resources, would need cognitive management to acquire significant, reliably formed beliefs. The debate over the importance of reliable formation to epistemic status, however, has focused on scenarios in which management is easy. There are possible processes that can be blindly trusted and still yield a great enough proportion of true beliefs in relevant environments to reliably generate true beliefs. They are the sort of processes found in thought-experiments of ungettiered, undefeated reliable belief that is intuitively not knowledge that we examined in §II.2.

For instance, the author of a guide to the *I Ching*, the ancient Chinese method of divination, wrote

> Someone once asked me if I did not worry about being too dependent on the I Ching. On consulting it, it replied 'If you had a good friend who knew the secrets of the kingdom and was able to help you in your work, wouldn't it be a shame not to make use of that friend.' (Anthony 1988)

Now suppose that consulting the *I Ching* is in fact a reliable method of divination. Even if it is, we would not want to say that you can get knowledge from a source if that source's deliverances are the only reason you have for trusting it. It should be easy to see that the problem is a lack of metacognitive control. If the *I Ching* were unreliable, the author would still use it to acquire beliefs because of the unreliably generated belief, acquired from the *I Ching*, that the *I Ching* is to be trusted. The agent lacks the sort of control that we exhibit with our usual ways of thinking. Thus, the person's beliefs acquired from the *I Ching* are not knowledge even if they are reliably formed. Very loosely, we could say that the problem is that the author does not track the reliability of the *I Ching*; that is, her attitude towards it is insensitive to its actual reliability or unreliability.

A more interesting, more complex case is BonJour's reliable clairvoyant Norman, which I have mentioned on several occasions. It will be worthwhile to look at it in detail here. Let me start by quoting BonJour's own description of the case:

> Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact, the belief is true and results from his clairvoyant power, under circumstances in which it is completely reliable (1980: 62).

BonJour further argues that it makes no difference whether Norman believes that he is clairvoyant, since any belief to that effect would be unjustified.

It seems that the germ of BonJour's objection to reliabilism is that Norman's clairvoyance does not require selective application. Norman's clairvoyance can only yield a significant proportion of falsehoods under conditions rare enough or different enough from his normal situation that he is reliably right in believing what his clairvoyance indicates even without taking account of those possibilities. If Norman did have the capacity to selective apply his clairvoyance, there would be a sense in which he did have evidence for its deliverances. He would have control processes that monitored the situation and were able to determine more or less when the clairvoyance could and could not be trusted. These control processes would monitor the clairvoyance and its outputs, relevant bits of the environment, background knowledge, other processes, and the like, would be able to determine more or less how to reliably form beliefs using clairvoyance, and would then be able to influence his underlying cognition in accordance with what it determined. The resulting picture makes the monitoring states look very much like evidence; the control processes, like forming beliefs on the basis of the available evidence.

This presumably would not count as basing belief on evidence on BonJour's conception thereof—most importantly, since we have no reason to think that the metacognitive monitoring is consciously accessible. But it is certainly close enough to believing on evidence that it drains his objection of any real plausibility. If his complaint is just that reliabilism does not guarantee reflective access to the grounds of one's beliefs, then we have already seen how to account for it. In §I.3, I proposed that conscious access to grounds adds value to reliably formed belief; in §II.2, we saw that it is not possible to account for our practice of epistemic evaluation by requiring that all epistemic desiderata be consciously accessible.

So let us suppose that Norman need not selectively apply his clairvoyance; that is, he can simply trust whatever it seems to tell him and the resulting belief will be reliably formed. Given this, it might be easy for him for resolve conflicts between it and other faculties. Information arising from his clairvoyance might just "trump" all other considerations. This is a reliable way of getting beliefs, but can hardly be considered rational. It would mean, for instance, that if Norman were in Washington shaking the hand of a man who looked exactly like the

President, he would have to believe that the President was in New York if his clairvoyance told him so, and even if he believed that he should not believe these promptings. This is neurosis, not intellectual virtue.

Of course, ordinary unreflective agents do not believe every prompting they receive. Not many people would believe their eyes if the President appeared in a golden spaceship with two heads and three arms. To say just that Norman has no "evidence" for the deliverances of his clairvoyance, or that he has failed to perform the "epistemic duty...to reflect critically upon one's beliefs" (1980: 63) leaves the case crucially ambiguous. It fails to distinguish between the case of an ordinary unreflective person, who hasn't really thought about when perception is misleading but knows it when he sees it, and the case of someone who wouldn't know a misperception if it seemed to bite him on the ass.

There is a certain lack of agreement in just how to read the Norman case. While it is generally agreed that there is some sense in which his belief-formation is less than ideal, some authors[1] maintain that Norman has knowledge nonetheless. The ambiguity I described in the last paragraph appears to be the source of the differences in intuitions. Suppose Norman could tell more or less when his clairvoyance is misleading, but his clairvoyance so rarely goes wrong that he never has to engage this capacity. Then we would be right to assimilate his clairvoyance to unreflective perception and memory. As I noted above, it makes no real difference if Norman does not actually exert any influence on his clairvoyant powers. What matters for control is that he could if he had to.

One imagines that what BonJour intends is more like the second possibility, where Norman has no capacity at all to determine if his clairvoyance is inaccurate. In *that* case, it's clear that Norman *is* being irresponsibly credulous. It would seem that even if his clairvoyance were to indicate something incredibly implausible by the standards of his own belief-set and inferential capacities, he would still believe it. We do not have to suppose that Norman is using a trumping

[1] See, e.g., Sosa (1991), as we saw in ch. IV; Bach (1985), and Bernecker (forthcoming).

rule to reach the conclusion that he lacks responsibility. Suppose, for instance, that conflicts never arise because his clairvoyance generates beliefs that his other processes cannot corroborate or controvert. (Suppose it only yields beliefs about the weather on Jupiter.) Even then, it would be clearly irresponsible to accept beliefs that cannot be corroborated by any other means.

If Norman lacks control, he will have other disabilities as well. Most importantly, without control Norman cannot use his clairvoyance to explore or inquire about the world. We can use our visual processes to explore the world and acquire significant beliefs, but this depends on our abilities to manipulate the inputs to vision so that we can see what we want to know. If I want to know what time it is, I can position my eyes an appropriate distance from the face of my watch, in good light, and thus form an RF belief. As I noted above, such behaviours exert indirect metacognitive control over one's processes; *ex hypothesi*, Norman could not engage in them. If what he *really* wants to know is not where the President is, but whether his financial advisor is in Brazil, he's out of luck. Even if these limitations aren't sufficient to deprive Norman of knowledge, they are certainly part of what makes the difference between thermometer-like reliability and real intellectual virtue.

Thus it seems that the trouble with BonJour's clairvoyants is that they lack control. In the next section I will refine the definition of intellectual virtue to take account of easy management cases.

Virtue refined

If Norman's clairvoyance can reliably produce true beliefs without having to be selectively applied, this means that belief-forming process types involving his clairvoyance (and possibly other processes, e.g., irresponsible methods of belief-fixation) yield a high ratio of true beliefs in normal environments. So it seems that there are two ways that we could require Norman to have control over his clairvoyance despite its reliability.

First, we could require that he be able to avoid trusting his clairvoyance in circumstances in which it would be misleading, even if these are distant enough from his usual situation not to constitute normal environments. (They might not be too distant, though, but only rare enough that even if Norman is misled in those situations, the frequency of true beliefs generated by his clairvoyance is still high enough for reliability.) Suppose aluminum-foil hats scramble clairvoyant signals, leading to deranged intuitions; and suppose that Norman, being stylish, would normally never wear such a thing. Then one sign that Norman can appropriately control his clairvoyance is that he would avoid believing the results of the scrambled signals even in the highly unusual circumstances in which he might end up wearing an aluminum-foil hat.

Intuitively, Norman is lucky to have reliable clairvoyant powers; it seems that he might easily have had an untrustworthy clairvoyance that would have led him to form false beliefs. When we measure reliability, of course, we hold the belief-forming process type constant, and cannot account for such intuitions. But the second way of ensuring he has control over his clairvoyance is to allow the process type to vary a bit, to see if Norman would still form a high ratio of true beliefs if his clairvoyance was slightly altered, or if it were a similar process that he might have had. For instance, suppose Norman's clairvoyance module was implanted by the CIA in a very tricky clandestine operation. If Norman could avoid forming false beliefs if he had had a botched surgery or (say) a Windows-based clairvoyance device, then he would seem to have appropriate control over the properly functioning device he has now.

In a particular case, it seems fairest to apply whichever possibility is more plausible; i.e., whichever involves closer possible worlds. These considerations give us the following rather complicated account of virtue. It *seems* to capture the intuitions here, although it is almost certain that it will have to be further refined before really doing the trick.

V is an intellectual virtue of a subject S iff:

(a) V is a stable capacity of S's to exert control over his processes in a way that allows S to form true beliefs and avoid forming false beliefs, and

(b) for all the processes under V's control, whichever of (i) or (ii) adverts to nearer possible worlds is satisfied:

  (i) the process could be prevented from generating false beliefs in a large proportion of the nearest possible worlds in which it otherwise would;

  (ii) the most similar possible processes that S might have had while remaining the same agent would not cause him to generate unreliably formed beliefs.

Condition (i) does not require total control over a process, since it refers only to performance in nearby worlds. In condition (ii), the idea is that making changes in the underlying process would not tend to lead the agent to form false beliefs, because the appropriate control mechanisms could compensate for the changes.

When I gave the preliminary definition of virtue in the last section, I was concerned with the control of processes that, like ordinary human ones, have to be selectively applied in order to lead to RF beliefs. If an agent has the capacity to satisfy (a) for such a process, he can satisfy (c). The reason is that the nearest possible worlds in which such a process might lead to false beliefs are ones that are relevant to determining whether the overall belief-forming process is reliable. Hence, if the agent can selectively apply the process, he satisfies (i).

Furthermore, if V is to be a stable capacity, it should not be easily misled by possible alterations in the processes it controls; this implies that either (ii) is satisfied or at least that (i) adverts to nearer possible worlds than (ii). Thus the processes for which my original definition was intended satisfy (c). The new definition extends the scope of the previous one without making it more restrictive; it makes the requirement of control consistent for all sorts of processes.

Epistemic virtue and subject-attributable desiderata

A subject with an epistemic virtue has a capacity to control his own belief-formation so that it yields reliably formed beliefs, so that their reliability proceeds from his own cognitive character. It is thus very plausible that we can identify subject-attributable reliability with believing out of epistemic virtue.

According to the credit theory of knowledge (see §II.3), a belief is knowledge iff its being true rather than false is attributable to the agent's own activities and capacities; i.e., knowledge is subject-attributable true belief. This is a stronger notion than SA reliability, and it does not necessarily follow from believing out of epistemic virtue. In a Gettier case, for instance, it appears that the fact that the agent's belief is reliable is attributable to him, but luck is importantly involved in his belief's being true rather than false. Suppose Smith drives past a barn at dusk and believes that he sees a barn. However, he does not know that most of what appear to be barns in this area are just façades, and this is one of the very few real barns. Smith's belief is acquired through a visual process that in typical human subjects is under effective metacognitive control, so he believes out of epistemic virtue. Since barn façades are on the whole very rare in Smith's normal environment, his belief is reliably formed. Given his metacognitive control over the belief-forming process, it seems that its reliability is attributable to him. However, his belief is true only because he had the good fortune to see one of the few real barns instead of a façade. Thus he lacks knowledge.

A thorough analysis of Gettier cases, lottery problems, and other sorts of situations in which agents do not receive credit for having a true belief would take us too far afield. Thus an analysis of subject-attributable truth, and how epistemic virtue contributes to it, will have to await another occasion.

However, it seems that except for the rare exceptions of Gettier cases, lottery paradoxes, and the like, virtuously formed true belief should be knowledge. Statistically speaking, reliably formed belief is usually true; similarly, subject-attributability reliability usually means subject-attributable truth. Thus we can use intuitions about knowledge to inform our understanding of epistemic virtue. This is very useful; while our intuitions about knowledge are fairly clear and determinate, notions like reliability and virtue are close to being technical terms in contemporary epistemology. (I will examine the notion of reliability in detail in the next chapter.) Hence, my arguments in the rest of this chapter will be driven by intuitions about knowledge. The notion of virtuously formed true belief is close enough to knowledge that this will prove unproblematic.

2. The subjective status of virtue

It is widely thought that to be known, beliefs must be well formed internally as well as externally. The intuition is that knowledge must have some sort of special subjective status; as well as being well formed in an external, truth-conducive sense, it must be right from one's own point of view. It is also sometimes thought that reliable processes integrated with cognitive character cannot by themselves confer this sort of status on beliefs. Thus, we should examine whether our account of virtue will require some other condition besides virtuous production to guarantee the right sort of subjective status.

The idea that knowledge must be subjectively as well as objectively appropriate presumably arises from the traditional view that knowledge requires justification. When Greco, Zagzebski, and Sosa discuss subjective conditions for knowledge, they call what they are after "justification". Putting the issue in terms of what sort of justification is necessary for knowledge is not helpful. What we are concerned with here is just a purportedly necessary condition for knowledge;

and necessary conditions for knowledge may, like ungettiered belief, play no other role in epistemic evaluation. To keep everything precise, let us call what we are after in this chapter "subjective aptness", the sort of internal well-formedness that is necessary for knowledge.

It is quite common to hold that some sort of subjective aptness is necessary for knowledge, although I will only consider in any detail virtue theorists' attempts to capture the intuition. Greco says, "it would seem that knowledge has to be subjectively appropriate as well as objectively reliable"; knowledge must "be well formed from the knower's point of view", and knowers must not only be reliable, but "be sensitive to their own reliability" (2000a: 180). Furthermore, virtuous production does not entail subjective appropriateness; some other condition must be added to the analysis of knowledge. Greco is very concerned to avoid adopting an overly strong requirement for subjective aptness (which drives his arguments against Sosa's perspectivism); he finally settles on grounding beliefs in countenanced dispositions as the appropriate sense of subjective justification.

Zagzebski (1996: 271) heartily endorses Kornblith's claim that "knowledge requires...belief which is arrived at in a subjectively correct manner". Beliefs are subjectively correct if they arise from veritistic motivations, since the motivation is what makes someone praiseworthy for believing correctly (243). Since veritistic motivations are necessary for virtue, this does not require a separate account of subjective aptness.

However, as we saw in §V.1 we cannot require veritistic motivations for knowledge. This account of subjective aptness is too strong. If we want to preserve a link between subjective aptness and motivations, we could fall back on Greco's account. It requires that agents form beliefs in the same way that they do when they are properly motivated, but not that they be properly motivated at the time. Thus, it avoids the counterexamples we saw in §V.1.

Sosa's view is somewhat more complex, because he thinks that there are two forms of knowledge, animal and reflective. Only the latter requires subjective aptness, which is determined by perspectival coherence; the former is mere virtuous production. But this sort of subjective status requires foundations that can only be animal knowledge (1991: 290). Reflective knowledge is, however, the higher intellectual good; not only is animal knowledge "a lesser grade of knowledge", but may even only be knowledge in a "metaphorical" sense (1991: 275).

This distinction between types of knowledge suggests a way of accommodating subjectivist intuitions. Perspectival coherence adds value to belief; thus beliefs that are known and that also have that special subjective status are better than ones that are not. Consider, for example, the difference between knowing that observed phenomenon o was caused by a Higgs boson and knowing that n is the 468th number in the Medicine Hat phone book. The former belief is, of course, far more valuable than the latter, but it is not any more *known*. Thus we can grant that if Stanley knew where he acquired his belief that the Battle of Hastings occurred in 1066 and that this source was reliable, he would be in a better position than he is not having access to those facts; but even without that access, he knows when the Battle of Hastings occurred.

So we can accommodate the intuitions by saying that the sort of subjective aptness that is necessary for knowledge is quite minimal, but the result of only satisfying this minimal requirement is knowledge that is of less importance and value than knowledge plus high-grade subjective status. It is not incoherent to say that S knows that p but could be in a much better position with respect to p. It is better to have access to compelling reasons for p than it is to simply trust an authority on the matter. For instance, it is better to know the proof that Zorn's Lemma is equivalent to the axiom of choice than simply to know that one's set theory instructor says so and can be trusted on the matter. But this does not mean that one cannot know by testimony that the two are equivalent.

In §II.3, I argued that there was a meaningful difference in value between attaining a true rather than false belief through one's own efforts and capacities and attaining a true belief only through happenstance. It is quite plausible to suppose that the former value is knowledge. In the past few chapters, I have been working with the seemingly plausible hypothesis that virtues are capacities that allow agents to make intellectual attainments in their own right. These two considerations lend force to the supposition that virtuously produced true belief is knowledge (except in Gettier cases and the like). That supposition, however, would imply that virtuously produced beliefs are *thereby* subjectively apt.

There is a distinct subjective status associated with the products of processes that are under effective control. The agent is sensitive to the origins of a virtuously produced belief, in that he is able to monitor the processes that engendered it and control their activity so that they reliably produce true beliefs and help him attain his other epistemic goals. Of course, he may not have conscious access to or beliefs about the etiology of his opinions. This should not be taken to preclude his belief's having a special status from not just his conscious or doxastic point of view, but from the point of view of his cognitive character as a whole—which goes beyond what he believes and what is available to introspection. Although this status is weak, it reflects the demands of realism; anything stronger is liable to be too strong.

It may be fruitful here to compare the status of a belief produced by controlled processes with Greco's proposal that subjective aptness is making use of countenanced dispositions. Greco argues that a belief formed from a countenanced disposition is well formed from one's own point of view because it arises from the stable aspects of character that one generally applies when motivated to believe the truth. The belief's etiology thus falls within the scope of what one would consider one's best behaviour (so to speak), and such an etiology, Greco maintains, should confer subjective aptness. Furthermore, agents are sensitive to the grounds of beliefs arising from countenanced dispositions in the sense that forming those beliefs from those grounds is something one would do when on one's best behaviour. As we saw in §IV.1, they are not, however, guaranteed to know when they are using countenanced dispositions.

The subjective status conferred by controlled processes is stronger than that Greco advances. The beliefs arise not just from dispositions exhibited when on one's best behaviour, but from dispositions to manipulate one's processes to attain the goals at hand. Likewise, the sensitivity to grounds involves an ability to refrain from believing on untrustworthy grounds and an ability to apply faculties to attain grounds appropriate for settling questions of interest. As with Greco's account of responsibility, effective control is not always consciously accessible; but as we have seen this is to the proposal's advantage. Moreover, we have already seen that the canonical examples of agents who have reliably formed but not subjectively apt beliefs—BonJour's clairvoyants, Plantinga's tumour victim, etc.—are also cases where agents lack effective control.

Since effective control is required for virtue, this proposal leads immediately to the result that virtuously formed belief is subjectively apt. This is exactly what one would expect. If a belief arises from one's cognitive character, it should have the appropriate subjective status; it should not have to satisfy other conditions to be appropriately integrated with the *subject*. Virtue theorists do not agree on subjective justification any more than epistemologists in general do. Sosa's perspectivism takes justification to be essentially reflective; Zagzebski, a matter of appropriate foundations in one's ethical (as well as epistemic) character; Greco, a matter of conformity with one's opinion of one's best behaviour. This is a particular case of the situation we examined in §I.1, where what is seemingly a debate over justification is really a debate over which aspect of an agent's character is most central to epistemic status. On the view that I am developing here, the answer is—whatever in the agent's character gives her a well-integrated system that attains cognitive contact with reality. Different aspects of character work together to engender an intellectually virtuous agent; knowledge must be tied to the whole, but not necessarily to any particular part.

3. Metaknowledge

I will conclude this chapter by looking at the status of metaknowledge on this view of intellectual virtue. First, we should consider the relationship between metacognitive monitoring and one's knowledge of one's own faculties. Monitoring processes does not imply having knowledge about them, which avoids regress problems. I will also briefly look at how this view treats bootstrapping, a highly controversial putative method for acquiring metaknowledge. Bootstrapping is interesting to us chiefly because it provides a useful illustration of the power of this approach for dealing with longstanding epistemological problems.

The status of metaknowledge

As I noted above, monitoring does not necessarily involve forming beliefs about the reliability of underlying processes. Whatever form monitoring states take, they are important for evaluating object beliefs only inasmuch as they contribute to control. Thus the status of the object beliefs is not affected by whether the metabeliefs involved in controlling them are themselves virtuously produced. This is quite important in preventing a potential infinite regress. If the states that monitored one's knowledge also had to be known (or, say, justified), then the monitoring processes would have to be under effective metacognitive control; but then the processes that monitored them would have to be monitored, and so on *ad infinitum*.

Metacognitive processes are of course cognitive processes, and like most of the rest, they are differentially reliable and need to be applied efficiently; those that generate beliefs also have the potential to conflict with other processes. Thus, sometimes, reliable metacognition may involve monitoring and controlling belief-forming processes. Here is an interesting example. Cornoldi (1998) reports that persons who take memory courses tend not to use the mnemonic techniques they learn there. They tend, he says, to underestimate the difficulty of memorial tasks and overestimate the difficulty of applying mnemonics; given these beliefs, using mnemonics seems to be too much trouble. This suggests that these persons need some meta-meta-cognitive control to make their estimates of problem difficulty more accurate.

Nonetheless, meta-meta-cognition is necessary for *object-level* knowledge only inasmuch as it makes for more accurate monitoring of object-level processes and thus for more effective control. It has no other value for the object-level belief. There is no regress because metacognitive processes are valuable only for their contribution to the agent's capacities, rather than as belief-forming processes themselves. Anything goes, as long as they can fill the function they are there to perform.

The situation is different when we look at metaknowledge. To see this, we will have to take a brief digression into this category of knowledge. It includes knowledge of one's processes, including their reliability in different circumstances; knowledge of the origins of one's beliefs; and knowledge of the epistemic status of one's beliefs. There is a tendency in the literature to propose higher standards for knowing about one's own cognition than for knowing about the rest of the world. As we have already seen, Sosa evaluates metabeliefs by their contribution to perspectival coherence, rather than by their being virtuously produced. Since coherence-seeking reason is a virtue, the standards for metaknowledge are effectively the same as those for reflective knowledge; i.e., perspectival coherence plus virtuous production. Interestingly, Sosa does not appear to allow for animal metaknowledge. I suspect that he thinks that the value of knowing about one's own mind arises just from its contribution to the higher state of reflective knowledge. As I argued above, coherence and having a perspective on the grounds of one's beliefs are both epistemic desiderata. But this implies that metaknowledge that contributes to reflective understanding is more valuable than metaknowledge that does not; it does not imply that the latter is not metaknowledge at all.

The tendency to impose stronger conditions for metaknowledge than for knowledge is found elsewhere as well. Zalabardo (forthcoming) notes that both Alston and van Cleve maintain that one must have evidence to know that a process is reliable, even though they do not think evidence is generally necessary for knowledge. Ian Evans, in an unpublished manuscript,[2] notes that many epistemologists take knowing that one knows to entail knowing that the conditions for knowledge are satisfied. For instance, Lehrer says that "[f]or S to know that S knows that p, S must know that the four conditions for knowing that p are all satisfied" (1974: 229), and Danto holds that "a correct theory of knowledge" is necessary for knowing that one knows (1967: 52). Even Goldman holds that knowing that one knows requires knowing that one uses reliable processes of belief formation (1986: 56-7). All these claims are suspect. If virtuous, ungettiered, etc. belief is sufficient for every other sort of knowledge, why shouldn't it be sufficient for metaknowledge as well? Lehrer and Danto's view is preposterous if generalized to other concepts: shall we suppose that a correct theory of causation is required to know that placing one's finger on a hot burner causes pain?

Let us examine knowing that one knows that p in detail; the results can be generalized to other sorts of metaknowledge. One cannot know that one knows that p just on the grounds that one believes that p. Too many of our beliefs fail to be knowledge for this procedure to be reliable. It seems that the metabelief must be grounded in something about the belief that distinguishes it from others that are not known. If one knows that p satisfies the conditions for knowledge, has access to the grounds for it, or one knows that it was RF, one can readily know that one knows that p. However, none of these is necessary for knowing that one knows. As long as there is some reliable indicator of knowledge, the belief based on it can be virtuously produced.

[2] "Knowing that One Knows Revisited", presented at the 2006 Pacific APA meeting.

For instance, when I grind coffee I know when it has reached the fineness I prefer. I usually know when I know this, because I know that as long as I am sufficiently attentive it is very rare for me to grind the beans too much or too little, and I can tell whether I've been paying attention. I thought I determined the fineness by watching how the grounds spin in the grinder until I recently discovered that I can grind coffee just as well in dim light. Furthermore, I haven't been able to tell exactly how I know. If I try to observe myself grinding the beans, I don't reliably grind them right. So I don't know the grounds of my knowledge, though I know my knowledge of the grounds.

It would be more valuable if I knew what process generates this knowledge, the conditions under which that process is reliable, and so forth. And of course I have no arguments by which to refute skepticism about my coffee-grinding skill. My little bit of metaknowledge makes rational a certain degree of confidence in my coffee-grinding ability, but makes no other contribution to my intellectual life. We should aim to understand our abilities—to know what we can do, what we cannot, and how we can improve and extend our contact with the world. Nonetheless, the possibility of more valuable states I could have with regard to my knowledge does not imply that I do not know that I know, any more than the banality of "My Boyfriend's Back" implies that I do not know the lyrics.

For these reasons, we should take metaknowledge to be distinguished by its content, not its standards: it is virtuously produced true belief about one's faculties, beliefs, or the epistemic status thereof. Among virtues for acquiring metaknowledge, some will be more important than others because they contribute to *significant* metaknowledge: knowledge about our faculties that we can use to further our epistemic goals.

Let us now return to the possible regress I defused above. Being indicative of truth, metacognitive monitoring seems like it would be a fruitful source of metaknowledge (when it is available for belief-formation, which is not entailed by its being available to control processes). While object knowledge does not require metacognitive control of monitoring and control processes, *if* metaknowledge is based on monitoring, to be virtuously produced it will. Thus knowing does not entail knowing that one knows, since the latter requires capacities not necessary for the former.

For instance, suppose H is a shortcut that for a certain domain approximates A, an algorithm that is too often intractable to be relied on in the field. Edgar has some experience with approximating A with H. He has learned to identify a class C of problems, and he uses H only to get solutions to problems in class C. Now suppose that H is not actually a good approximation of A throughout C; but, the problems in C for which H does not approximate A arise *very* rarely in practice.

Barring fourth-condition complications, when Edgar applies H to a problem in C, he knows approximately[3] what answer A would yield. His belief is reliably formed, and this reliability proceeds from his own capacities and character—namely, his ability to identify the right problems to which to apply H. Now suppose he believes that he knows, and this belief is based on his ability to identify problems in class C. This belief is a dubious candidate for knowledge, because the link between class C and the reliability of H is tenuous, depending only on a fortuitous environment. If Edgar does not know the facts about H's behaviour in class C, it seems that he cannot claim to know that he knows.

Put in the terms of the analysis of virtue I gave above, Edgar can control his application of H so that it allows him to reliably form beliefs, and thus his beliefs about approximately what answer A would yield are virtuously produced. Since he cannot distinguish the problems in class C for which H approximates A from the ones in which it does not, he lacks appropriate control over his ability to monitor H's reliability. Thus he does not know that he knows; his metabeliefs are not grounded in adequate cognitive management.

[3] "Knows approximately" shouldn't be thought problematic; given the limits on the precision with which we can—or bother to—make measurements, much of our knowledge takes this form. There is certainly much to be said about knowledge involving vague terms (a study inaugurated by Timothy Williamson), but here is not the place to say it.

Bootstrapping

Bootstrapping is a putative method for acquiring metaknowledge that has excited a great deal of controversy. It may provide a useful illustration of the principles for which I am arguing to see how we can analyze bootstrapping. The procedure starts by adducing premises of the form

Process P generates the belief that p_i, and p_i is true.

With enough such premises, the agent can reason inductively to the conclusion that P is reliable. The problem is that this argument places no restriction on how each p_i is known. Each one may just have been formed by process P, and so the agent may have no other reason to believe them other than their being generated by P. Even so, if P is reliable, the conclusion is reliably formed (see Zalabardo forthcoming), since it is an inductive inference from true premises.

On the one hand, bootstrapping is a truth-preserving inference. Furthermore, it makes for easy responses to skepticism. It allows us, for instance, to infer from our past inductive successes that induction is reliable (van Cleve 1984). On the other hand, it has an air of circularity to it. It is highly counterintuitive that one can learn that a process is reliable just by reasoning from the deliverances of that process. Now this is sometimes plausible for processes that, like induction or sensory perception, are so general or so basic that it is hard to see how we could test their performance against something that does not presuppose them in some way. Thus bootstrapping can seem a promising approach to skepticism. But that still doesn't make it a very plausible way of getting knowledge about one's processes more generally.

The fate of bootstrapping is usually tied to that of reliabilism. It is often argued that bootstrapping appears circular because in order to know that p_i based on its being generated by process P, one must already have warrant for, i.e., be in a position to know that, P is reliable. Thus bootstrapping does not provide any additional support for the conclusion, though perhaps it might make explicit the status the belief already has. (See Zalabardo forthcoming, Cohen 2002.) The trouble is, of course, that such restrictions on knowledge seem to require the sort of access to the grounds of one's beliefs that we found problematic in §II.2. So the problem is to find a suitable way of ruling out bootstrapping that is not more counterintuitive than its intended target.

Rather than get too deep into the controversy, let us look at how bootstrapping fares on the theory of virtue presented here. Bootstrapping is yet another example of a differentially reliable process; it is a truth-preserving way of acquiring beliefs about reliable processes, but only when the process described in the premises is actually reliable.[4] Hence, a virtuous faculty of bootstrapping requires a capacity to make the inductive inference described plus a capacity to identify which processes may be bootstrapped and which may not.[5]

On this view of knowledge the trouble with bootstrapping is that it yields knowledge that a process is reliable only if the agent can already identify reliable processes to which to apply it. This conclusion is similar in some ways to the response that in order to know the premises of the bootstrapping argument, the agent must already be in a position to know the conclusion. But as we saw above, metacognitive control over P does not imply being in a position to know that P is reliable. Rather, what it comes down to is whether the agent has enough control over the application of induction to her own capacities to be able to avoid trying to bootstrap unreliable processes. If she does, bootstrapping consists of a move from this capacity to pick out reliable processes to virtuously formed beliefs that those processes are reliable.

Then when the agent can correctly apply it, bootstrapping *is* a way of

[4] If we were to treat bootstrapping as a very broad process type, it would be unreliable, since it concludes that the object process is reliable whether applied to a reliable process or not. This doesn't help the reliabilist very much. As we'll see in the next chapter, reliabilism generally requires very narrow process individuations. We have no reason to think bootstrapping should be individuated broadly, except that it would be very convenient if it were. (Cp. Vogel 2000.)

[5] As well, the agent has to recognize that all the beliefs mentioned in the premises have a common source. This is not always a trivial step.

getting knowledge about reliable processes. After all, inductive reasoning in general requires selective application. It seems that most introductions to induction use "the sun will rise tomorrow" as an illustration of a sentence that is confirmed by one's past experience. The example only works because very few readers of such books spend their winters north of the Arctic Circle. Our inductive reasoning is generally reliable, of course, because we have a variety of habits and theories of varying levels of explicitness that we can use to decide whether induction can be applied to any particular set of phenomena.

Perhaps bootstrapping has its air of illegitimacy because the capacity to choose a right set of premises is what does the real work. This isn't generally true of inductive inference; a background theory can tell us that Φ is projectible but not tell us whether all Ψs are Φ. But it's a plausible hypothesis that if one had sufficient insight into the grounds of one's capacities to identify reliable processes, then that would be adequate grounds for knowing the process to be reliable. For instance, suppose S implicitly uses the coherence of a process's products as a guide to whether bootstrapping may be applied. She is unaware of doing so; to her, some processes (the ones that cohere well) just seem bootstrappable. Then it seems that if she can know P is reliable through bootstrapping, she could also know it by reasoning

> Beliefs formed by P cohere with the rest of my beliefs;
> therefore, P is reliable.

Or, the capacity to identify reliable processes could underlie a directly formed belief in the reliability of that process without a detour into inductive reasoning. One might just see that P is reliable, making use of the same capacity that would allow one to identify premises for bootstrapping.

These are, of course, substantive claims about capacities to identify reliable processes that we need not investigate in detail here. Nonetheless, they make it *prima facie* plausible that bootstrapping can yield knowledge under the right circumstances, but that in those circumstances it could just as easily be replaced by other methods of acquiring the same knowledge. Bootstrapping would be essentially superfluous as a source of metaknowledge. But this is not because knowing the premises entails being in a position to know the conclusion. Rather, it is because being able to identify appropriate premises for bootstrapping arguments entails having other ways of getting the same metaknowledge.

VIII

RELIABILITY AND GENERALITY

In this chapter, I will look at two closely related problems. The first is the well-known *generality problem.* Reliability is a property of a belief-forming process type, but any belief arises from a process token that can instantiate many types. Thus, reliabilists need to establish that there is a fact of the matter about what process type generates a belief. The second problem is unique to the position that I have developed here; it consists of finding determinate relationships of control among processes. If there is no fact of the matter about whether a process is under effective control, there is no fact of the matter as to whether the agent has intellectual virtues.

What I will do in this chapter is try to make it plausible that these two problems can be solved. Roughly, the answers to both are determined by facts about the structure of the cognitive system in question. To do this, I will tentatively advance what seems to be a promising account of the generality problem. Any solution along those lines must take account of the internal structure of belief-forming processes, including structures of monitoring and control. Thus being able to solve the generality problem means being able to individuate relationships of metacognitive control. The account of virtue I advanced above does not make us any worse off with respect to the problem of individuating processes.

The approach that I will take is a form of *process reliabilism.* Sosa and Greco have their own approaches to the generality problem, which I summarized in §IV.1. I won't examine these approaches in much detail here. Space does not permit a detailed analysis of different approaches to the generality problem. What I'm concerned with is showing that there is *some* way of making the analysis of virtue in the last chapter rigorous, by making rigorous the notion of process types. Whether it's the only way, or the best way, are not issues that I'll consider here. A full account of the generality problem would be a dissertation in itself.

The generality problem

Reliability is a function R(P, E, Φ), with P a belief-forming process (BFP) type, E a set of environments, and Φ a set of propositions. R is the propensity for instances of P to produce true beliefs on propositions in Φ in environments in E.

The theory of knowledge is mostly concerned with whether a process is sufficiently reliable for its deliverances to be known; this can be represented as R_k(P, E, Φ), which is true iff the probability R(P, E, Φ) exceeds some vague threshold k. The most we can say about the threshold is that it is high enough to allow us to hold reliably formed beliefs confidently. Reliably formed beliefs are fallible, but the possibility of error should be small enough that we can ignore it most of the time.

It would be otiose to try to be more precise than this about the location of the threshold. Like any other vague boundary line, *if* a precise boundary exists there is no principled way of determining where it is. When we need a precise threshold to work with, we can select an arbitrary point from the range of acceptable candidates. It is interesting to suppose that the threshold might vary with the cost of error.[1] It would certainly make sense to demand a higher threshold for knowing that the blowfish was prepared correctly than for knowing that the miso soup was. However, we won't worry about such complexities here.

Strictly speaking, reliability is a property of a process type. The reliability of a process is sometimes important in epistemic evaluation. For instance, when we decide what processes to use in inquiry or what authorities to trust, we consider how reliable they are. More important for epistemology is the notion of

[1] Which would incorporate some elements of Hawthorne's (2004) "moderate invariantism" into our theory of knowledge.

a belief's being reliably formed. This is a departure from ordinary language, in which "reliable" cannot be predicated of a single event. We have

(1) My car starts reliably in winter,

(2) Kobe Bryant reliably makes free throws,

but not

(3) *That free throw was made reliably,

(4) *The car started reliably just now.

The reason for this seems to be that sentences like (1) and (2) in context implicitly specify a process type and set of environments over which reliability is being measured. "My car" in (1) refers to my car in its normal condition; the range of environments is the sorts of Edmonton winters that we've had since I got my car. In (2), the process type is Kobe Bryant in the usual physical condition in which he plays a game, and the relevant set of environments are typical sorts of games. Thus these sentences have reasonably clear truth-conditions.

When we look at single events, however, what we have are process tokens that instantiate many types, and a context that underdetermines the range of relevant environments. The process type in (3) might be Kobe in the usual physical condition in which he plays a game, or Kobe in his usual physical condition for the end of the third quarter; we can go as specific as we like, down to specifying Kobe's condition molecule-for-molecule. The relevant environments might be as broad as typical sorts of free throws, or typical sorts of third quarters when the Lakers are down by 12, all the way to molecule-for-molecule replicas of the entire game so far. Sentences (3) and (4) scarcely hint as to how they are to be evaluated.

Thus the attempt to apply the concept of reliability to single events gives rise to the *generality problem*, the problem of going from a belief that p generated by process token t in environment token e to the process type P and classes of environments E and propositions Φ that determines the reliability of the tokens.[2] The problem is not so much that we need a feasible procedure for calculating the reliability of any given belief. Rather, the problem is that we have no assurance that there is a fact of the matter about the reliability of a token belief arising from a token process.

The generality problem is closely linked to the *reference class problem* of going from the propensity of things of type A being Bs to the probability that a token of type A will be a B. If we can identify the relevant P, E, and Φ for a belief that p, R(P, E, Φ) can be the objective epistemic probability of Bp (Alston 2005: ch. 5). However, since we are not concerned with epistemic probabilities here, we won't worry about this link. The response to the generality problem we will examine below is quite specific to belief-forming processes, and does not offer any substantial morals for the reference class problem.

In what follows, I will lay out an account of how to individuate cognitive processes that will solve the generality problem and allow us to individuate monitoring and control processes. We can call this the *trilevel* approach (for reasons that will be clear shortly); it is a form of process reliabilism.

Process reliabilism

Process reliabilism is marked by taking E to be held constant for each agent (at a time, broadly construed) and Φ to be the set of all propositions. We measure the performance of the process type P over a standard set of

[2] See Conee & Feldman 1998 for the canonical statement of the problem.

environments for all the propositions in which it generates beliefs.[3] Thus only the propositions in Φ in which P generates belief are relevant to its reliability. On this approach, what does the real work is how BFP types are individuated. Different beliefs have different statuses when they are generated by different types of BFPs. This is in contrast to Sosa's approach, which is to take the process type to be the same for all beliefs; roughly, it is determined by those facts about the agent's constitution that are stable across nearby possible worlds. Beliefs that have different statuses are associated with different E-Φ pairs. We looked briefly at the very difficult question of how E-Φ pairs are chosen in §IV.1.

On the most plausible sort of process reliabilism, E is determined by the environments that are typical for the agent in question in her ordinary life, containing a sufficiently wide range thereof in approximately the frequency in which they are apt to occur. Thus E ignores highly atypical situations—being at the centre of the sun (in a very good protective suit), or in barn façade country, or looking at a painted mule in a zoo. The upshot is that we require that processes be likely to yield truths in situations that are not particularly different from those the agent could expect to be in, and we rule out situations that the agent is very unlikely to find himself in. Greco (2000a: 211-7) notes that success over such situations is part of what we require for an agent to have an ability (in this case, an ability to form true beliefs).

As Alston notes, this suggestion

> is far from precise...[but] this suggestion has the right kind and degree of sloppiness for the concept of reliability we want for epistemic purposes. It does unequivocally rule out clearly atypical situations—Cartesian demons, brains in vats, and the like. And it makes a judgment of reliability dependent on our actual situation as human beings in the environments in which we actually find ourselves (1995: 10).

[3] Goldman (1986: 44-5) calls this *global* reliability; in *local* reliability, which he also requires for knowledge, Φ is restricted to the content of the belief being appraised. Local reliability entirely fails to avoid the typical-truth problem (see below); thus, it does not seem to play an important role in the appraisal of beliefs.

There is always a certain amount of vagueness about what counterfactual situations, especially what possibilities of error, are relevant to evaluating a belief. A vague specification of the environments over which we appraise reliability allows us to capture the vagueness in our intuitions (cp. Greco 2000a:15-6). Thus the vagueness of E is not a weakness of the analysis; at least as far as accounting for common usage goes, it is a strength.

Nonetheless, in identifying E and Φ this way we only defer the generality problem, since any process token instantiates many types that can vary in reliability over typical environments. I look at my desk right now and I see a blue pen. This process can be characterized as *perception, seeing, seeing in good light, seeing an object 2' away in good light, seeing what looks like a pen 2' away in good light, seeing what looks like a blue pen 2' away,* and so on. Or we might characterize it as responding to a particular sequence of sensory inputs that we might describe with varying levels of detail, from specifying the exact inputs I received, to more general characterizations, all the way up to *a cognitive process from inputs to beliefs*. Finally, there are many other properties we could include in our specification: it might be *a man with an unkempt beard seeing*, or *seeing on a Friday*, or *seeing while eating a nacho*. The reliability of these different process types can vary extensively. *Seeing a pen* is less reliable than *seeing a pen from 2' away*, but more reliable than *seeing something out of the corner of one's eye while thinking about how to phrase a sentence.*

There are several important constraints on plausible ways of individuating BFP types.[4] In general, we have to respect our intuitive appraisals of various beliefs. More specifically, we need to respect their fineness. If we individuate BFPs too broadly, we run into the *no-distinction problem.* If two beliefs are generated by the same BFP type, one is reliably formed iff the other is. The no-distinction problem arises when BFPs are individuated so broadly that the same one generates beliefs with different intuitive statuses. If we individuate processes

[4] And E-Φ class pairs on Sosa's approach.

too narrowly, we can run into the *typical-truth problem.* If P is narrow enough that it only produces belief in one proposition that is true in every environment in E in which P occurs, then P is automatically reliable. For instance, the process type *believing that one is not a brain in a vat because one received a D in Philosophy 102 and as a result hates speculative epistemology* is perfectly reliable, since we are not brains in vats in any typical environments. Two special cases of this problem are more widely discussed in the literature: the "necessary truths problem", in which the proposition true everywhere in E is necessary; and the "single-truth problem", in which a process that actually generated a true belief individuated so narrowly that it only occurs in the actual world and is thus reliable by default.

The trilevel condition

Several responses to the generality problem are similar enough that we can regard them as variants on a single approach, which we can call "psychological realism". The central idea is that belief-forming processes (BFPs) are real psychological functions, and part of the job of cognitive science is to taxonomize them. If functionalism or something like it is true, any token belief must result from a single function type. The most general statement of the view is in Alston (1995), but it has been developed further by a number of others, and the main points were previously articulated by Wallis (1994).

The first main point of the approach is that BFPs supervene on the cognitive system in which they occur. Thus their tokens cannot extend outside the cognitive system, and their types are individuated only by cognitive properties. As I argued in §V.2, this does not mean that the realization of a BFP cannot extend outside the organism's body (*contra* Alston 1995: 11). Cognitive processes can include external vehicles. Nonetheless, this is a very strict condition. It rules out types like *a man with an unkempt beard seeing*, or *seeing on a Friday*, since the condition of one's hair and the date are not cognitive properties.

I won't offer here a full account of what cognitive properties are or how to identify BFP tokens.[5] That is a problem for cognitive science, not epistemology. What is important is that besides raw pessimism, we have no reason to think that cognitive science lacks the resources for determining what properties are relevant to their field of study. A crucial step in this defense of reliabilism is showing that if a completed psychology is possible, reliabilism can succeed. Thus we reduce worries about BFP individuation to general worries about the determinacy of mental functions and our capacity to understand them. But if a general skepticism about the possibility of our understanding our own minds is right, the failure of reliabilism is the least of our problems.

The version of the psychological realist approach that I will give here is based on Beebe (2004). Marr's "tri-level hypothesis", which informs much of cognitive science, is central to his approach.[6] Marr proposed that cognitive behaviour must be explained at three levels: the *computational* level, or the information-processing problem being solved; the *algorithmic* level, or the particular algorithm that the system uses to solve the problem; and the *implementational* level, or how that algorithm is implemented in the physical system. All three levels involve different sorts of explanation and a different stable of explanatory concepts. Explanations at each level are necessary for understanding information processing, and no one level is explanatorily sufficient. The computational level is abstract and semantic. It tells what different states represent, what the agent is trying to do in the environment, and how to interpret different stages in the cognitive process. The algorithmic level describes the procedure being used, and gives a syntactic characterization of the psychological functions involved. The implementational level describes the physical system in which the cognitive process is realized, and how the algorithm is implemented in that system.

[5] See Adler & Levin (2002) and Beebe (2004) for discussion.

[6] See Dawson (1998) for a general discussion of Marr's hypothesis.

BFP types are "information-processing types" (Beebe 2004: 183), and in accordance with Marr's hypothesis, they are determined at all three levels: by the information-processing problem being solved, the algorithm being used to solve it, and the physical instantiation of the algorithm. Two tokens must instantiate the same type at each level to be the same BFP type. Beebe calls this the *tri-level* condition. Of course, these are *types* at all three levels—problem types, algorithm types, and physical-system types. Thus it may seem that we have only pushed back the generality problem. But part of the job of psychology is to identify kinds at these three levels. If there is no fact of the matter about, say, whether Bob and Rob use the same algorithm to multiply large numbers, we have problems that go far beyond the failure of reliabilism. Luckily, there is no reason to think that cognitive science lacks the resources to identify types at these different levels.

One reason the generality problem gets its intuitive grip is that we have a bad habit of informally thinking of BFPs at only one level at a time. Conee and Feldman, for instance, describe a sequence of neural events starting with retinal stimulation and ending in a belief in a nearby maple tree, and then declare, "[t]his sequence of concrete events is the process that caused the belief" (1998: 2). The trouble is that except for the mention that this is a process terminating in a belief, the description of the token occurs only at the implementational level. The physical properties of the token do not determine what type of BFP it is because BFP typehood is determined by computational and algorithmic properties as well (Beebe 2004: 184-7). At the other extreme, processes like *seeing in good light* or *seeing a pen* are little more than vague computational-level descriptions.

It is interesting that in commonsense epistemology we come closest to giving process descriptions that satisfy the tri-level condition when we describe the evidence on which a belief is based. Suppose we say that S believes that tree T is an elm because it appears to him that the leaves have nine notches on them, and he believes that only elms have leaves with nine notches. We have here a rough description of the algorithm S used, which consists of integrating information derived from visual representations of the tree's leaves with background information about elms. The description proceeds semantically rather than syntactically, which makes the description less rigorous but also gives a rough characterization of the process at the computational level. Information about the physical implementation of the algorithm is implied though not stated; we may assume that it is implemented in the usual human way. Any remaining difficulty in determining just how reliably formed the belief is arises from the vagueness of the skeletal description.

It is not particularly controversial that there is no principled difficulty in evaluating a belief on the truth-conduciveness of the evidence on which it is based. On the contrary, evidentialists (who include some of the most trenchant critics of process reliabilism) are committed to there being a fact of the matter as to how reliably formed a belief is, given that it is based on certain evidence (Comesaña 2006). Of course, they may not think that the proportion of true beliefs based on that evidence in nearby worlds is what determines the status of the belief. But if there are facts about the truth-conduciveness of evidence in nearby worlds, it is a short step to finding facts about how reliably formed such beliefs are. The trilevel approach to the generality problem supposes that the reason it is so easy to talk about the reliability of belief based on evidence is that those token-descriptions contain information at all three levels.

The role of metacognition

Nonetheless, we cannot just rely on information-processing types as identified by a completed psychology. Reliabilism needs to avoid the no-distinction problem; it needs to draw distinctions between processes that are fine enough to reflect the distinctions we draw between the epistemic statuses of different beliefs. Scientific explanations of psychological phenomena, however, are meant to be as broad ranging as possible, so that the most phenomena possible can be explained by the invocation of the smallest number of theoretical laws.

The result is a tension between the needs of psychological explanation and epistemic evaluation that makes it highly unlikely the two should identify types in exactly the same way.

Suppose, for instance, that it turned out that all human inductive reasoning could be treated as a single computational problem carried out by a single very complex algorithm. This algorithm would take a huge range of inputs; suppose that the huge number of different possible combinations of inputs explains all the different outputs that inductive reasoning can provide. If this were discovered (and I am not suggesting that it will be), it would be a major advance for cognitive science, since it would provide a single unified theory of a wide range of cognitive behaviour. But if reliabilists were beholden to describing types just as psychologists do, this discovery would be a complete disaster for them. If all inductive reasoning were the same information-processing problem, then all its outputs would be equally reliable.

Thus reliabilism can only work if we can get into the internal structure of information-processing types. When agents use broad-ranging information-processing types, we have to be able to identify relevant subprocesses within them in order to distinguish the etiologies of beliefs formed in relevantly different ways.

One important element of structure that we will need is to uncover control relations. Since psychological processes are real entities described by cognitive science, relationships of metacognitive control can be individuated in the same way as BFPs. We look for relationships of monitoring and control between processes, which should be apparent at the computational and algorithmic levels. In §VI.1 we looked at three distinguishing features of metacognitive processes. First, there is a distinction between metalevel processes and object level processes; second, the metalevel functions as a model of the object level; third, the metalevel regulates the object-level by initiating, sustaining, or terminating activity therein. The facts about whether processes have these features and solve those problems, and thus whether agents have appropriate control over their processes, are just as objective as the rest of the facts about the structure of the cognitive system.

Meta-level processes implicitly distinguish object-level processes by what they are able to differentially influence and what they are not. For instance, suppose a control process can initiate or terminate a memory search, but cannot have different effects on memory searches that retrieve a previously stored trace than on memory searches that reconstruct an ostensible memory out of nothing. Then the control process implicitly distinguishes memory searches as a type separate from others, but treats veridical and non-veridical reconstructions as the same type. Of course, in normal humans there are a number of ways of distinguishing ostensible from real memories. Thus another control process might differentiate, say, fluently retrieved data (which it allows to be believed) from that which is nonfluently retrieved.

When individuating BFP types, we must type controlled subprocesses in accordance with the combined effect of any control processes operative in forming the belief. This can probably be stated only vaguely in the general case; we would need to see just how different control processes influence belief-formation to determine what this means in particular cases. Nonetheless, it is necessary for handling what we might call "internal lucky evidence cases" like the following. Suppose Sam can't find his keys. He reasons that they probably fell out of his pants pocket when he was sitting with his feet up on his desk (um, thinking). It is highly unlikely that this would have happened, but Sam does not realize this. Then he happens to remember that he changed his clothes earlier and left his keys in the pocket of the pants he was wearing before. He could very easily have failed to remember this, and so could very easily have persisted in the false belief that his keys are on his office floor.

Very roughly, we can take the overall strategy involved here to be something like *check memory for information about keys; if none is found, use* *default reasoning*. Since the default reasoning would yield a false belief and the memory check might easily have failed, the reliability of the whole strategy is quite low. Intuitively, of course, given that the memory check did not fail, the possibility that it might have is no longer relevant to whether Sam's belief is reliably formed.

Analogously, suppose Stan is very likely to remember where he left his keys, but by chance on this occasion the memory check fails. Then the reliability of the overall strategy is quite high, but in evaluating his belief we consider only the (low) reliability of the default reasoning that actually formed it.

What allows us to type this process correctly is the fact that Sam and Stan's control capacities make an implicit distinction between the memory check and the default reasoning. This emerges in the fact that they give memory priority over default reasoning; they only allow the belief to be formed from default reasoning if the memory check fails to yield a suitable answer. The two subprocesses can be controlled differentially, in that the metaprocesses can have different effects on them, and these effects and the circumstances in which they differ are appropriate for effective regulation. When considering the final belief-forming type we respect this implicit distinction, and treat beliefs resulting from memory checks and beliefs resulting from default reasoning as arising from different process types. In general, where one's metacognitive processes implicitly distinguish p and p′, then for purposes of epistemic evaluation p and p′ are different process types.

On the other hand, suppose that Spam is in the same situation, and he resolves conflicts of this sort by believing whatever process terminates first. Thus he has no serious capacity for control. Intuitively, the reliability of Spam's BFP is determined by the reliability of the whole process. Whichever subprocess finishes first, the reliability of his belief is determined by that of the two subprocesses and the propensity of each to win the race. The reason is, of course, that Spam's processes do not implicitly distinguish the two subprocesses, and thus the whole process is what determines the reliability.

Another important implication of the role of control processes is that they reflect an implicit sort of awareness the cognitive system has about its own constituents. Sosa proposes that subjects whose habits of belief-formation are guided by representations of E and Φ pairs have "implicit beliefs" about those E and Φ pairs for which they are reliable and those for which they are not. (Recall that Sosa takes the BFP type to be the agent, and associates beliefs with different statuses with different E-Φ pairs.) The generality problem is solved by distinguishing the same E and Φ pairs that the agent's implicit and explicit beliefs do. (See §IV.1.) It is by no means clear if this proposal is determinate enough to work, given that it only takes account of agent's representations of their processes and not the processes themselves. Nonetheless, there is a *prima facie* plausibility to the general idea of individuating processes in the same way that the agent does (inasmuch as this is possible). The trilevel condition captures the distinctions that control processes are capable of making in the course of regulating belief-formation. Thus it does take account of the agent's own ways of implicitly distinguishing his own processes.

Conclusions

I would tentatively suggest that the above account provides an adequate solution to the generality problem. It is possible that we may need to take account of more factors when tracing out process types from descriptions of the structure of cognitive processes; nevertheless, it seems likely that the trilevel approach is on the right track. However, a complete defense of it will have to await another occasion.

More important are the lessons we drew from the cases of Sam, Stan, and Spam above. This illustrates that however we try to solve the generality problem, we must take account of relationships of metacognitive control. So if we can solve the first problem I described at the beginning of this chapter, the problem of finding determinate process types; we can solve the second, the problem of finding determinate relations of metacognitive control. This gives us one conclusion we can advance quite firmly: the account of virtue I have urged here is no more problematic than any other form of reliabilism. Any successful account of process individuation has to have the resources to identify metacognitive control. Thus, there is no special difficulty in taking it to underlie intellectual virtue.

IX

THE STRUCTURE OF INTELLECTUAL VIRTUE

In §VII.1 I defined an epistemic virtue as (roughly) a capacity to control one's processes so that they reliably yield true beliefs. This gives us an account of virtues that produce knowledge. In §III.1, however, we saw that "high-level" virtues do not generate knowledge. One can believe that p out of intellectual courage or open-mindedness (for instance) without knowing that p, and one cannot advert to those virtues when giving reasons for why one knows that p. A full account of intellectual virtues must be more general than I have heretofore provided.

The previous discussion, however, concerned only one epistemic good—reliability. However, the aim of belief-formation is not just to avoid error, but also to attain significant true beliefs in a sufficiently wide range of situations. Thus there are other valuable aspects of cognition besides reliability. Control capacities can be dedicated to increasing the range of true beliefs that can be acquired, the range of environments in which they are acquired, and so forth. As I argued in §II.3 and §V.1, when a desideratum arises from a virtue, its obtaining is attributable to the subject. This is the case even when a virtue increases the range or significance, rather than reliability, of its possessor's beliefs. Control capacities can even themselves be evaluated by the range of processes they control and environments in which they operate. Given all the ways that a process can be valuable or give rise to subject-attributable epistemic value, as I will argue in this chapter a generalization of the account in §VII.1 can capture a wide range of virtue-theoretic evaluations.

1. Process desiderata

In chh. I-II we examined desiderata of belief. Now we will look at desiderata of processes as well. I will discuss some of the more important of these, and say a bit about how metacognition can contribute to them. Although reliability is certainly a process desideratum, we have already discussed it extensively and I won't repeat myself here. Three other subsidiary goals of cognition will need brief discussions: power, portability, and significance. I will also examine the value that attaches to control capacities themselves when they exhibit these desiderata. All these considerations will lay the groundwork for the generalization of the definition of virtue that I will present in the next section.

Power

Power is the capacity to acquire or generate a large number of true beliefs. While reliability guards against error, power guards against ignorance, or the lack of true beliefs on important matters. It is at least a sliding scale, and it may be easiest to think of it as essentially comparative: P_1 is more powerful than P_2 iff the true outputs of P_2 are a proper subset of those of P_1. The resulting partial ordering is all we need for my purposes here.

Power is only worthwhile when combined with a certain degree of reliability. It is no good acquiring lots of new true beliefs if the cost is acquiring even more false ones. At the very least, then, when evaluating power we also need to consider the proportion of false beliefs acquired. Goldman proposes that we can evaluate power just by looking at a subject's performance over a given subject-matter and considering the proportion of true beliefs she forms versus false beliefs or failures to form a belief (1992: 167-8). We might speculate further that power is the capacity to acquire a large amount of knowledge. Although this is an intriguing possibility, I will not argue for it here. Let us just take power to be a capacity to acquire large numbers of truths without too many falsehoods.

Since the goal of cognition is true beliefs on matters of importance or interest, what is really valuable is the range of significant true beliefs a process can form. Increasing the range of square metres of the Sahara in which one can count the grains of sand does not make the processes involved any more valuable. Thus we should take power to be the range of significant true beliefs that a process generates, without also generating too many false ones.

Note that for practical purposes we can legitimately evaluate the range of true beliefs an agent can form even when they are not significant to him. I have only very weak capacities for forming true beliefs about corporate finances, because like most people I find accounting too dull to cultivate as an occupation or hobby, and I have little reason to form beliefs about the health of other people's companies. This incapacity does have consequences; it means, for instance, that no one should hire me to be their accountant. (While it might seem inconceivable that this would be relevant, a friend of mine did once try to persuade me to handle the accounting for a small business he was starting up.) Likewise, I have extremely weak tree-identification capacities, having grown up in the suburbs and never had an interest in the subject. Everything I know about what elms look like I learned from Conee & Feldman 1998. Among other things, this indicates that if you like to learn about the trees you see, you had better take someone else on your nature walks.

These are both legitimate evaluations of my cognitive capacities, though ones with only very narrow applicability—narrow, of course, because it says little about my cognitive capacities to observe that I lack the power to acquire beliefs I do not care to have. (Note for instance that it doesn't mean I couldn't acquire powerful accounting or tree-identification capacities if I had a good reason to.) Since such appraisals are so narrow, they have generally been ignored by epistemologists—a trend that I intend to continue here, aside from these brief remarks.

When we evaluate power, we normally construe processes more broadly than when we determine whether a belief is reliably formed. The power of one's tree-identification capacities, for instance, is (intuitively) a matter of the range of true beliefs one can form without too many errors over a range of arboreal species. But take an agent who knows that he is less reliable with identifying elms than identifying maples. Intuitively, his unreliability with elms does not detract from the reliability of his beliefs about the presence of maples. For purposes of appraising reliability, elm-identifications and maple-identifications arise from separate processes, but for purposes of appraising power, they arise from the same process.

Moreover, we often evaluate an agent's entire bundle of processes over a certain range of significant propositions (for instance, accounting or tree identification). The power of a single process is important for some purposes; for instance, when trying to decide how to conduct one's inquiries, it is important to initiate powerful processes (at least, powerful enough to yield true beliefs on the questions of interest). But having a weak process does not necessarily restrict the range of true beliefs an agent can acquire, because he might have other processes that can do the job. Noting only that the blind have no visual processes of any power at all exaggerates the extent of their disability. Their other perceptual processes are normally more powerful than average, especially with regard to true beliefs the sighted acquire through vision. Similarly, if I have a calculator handy I will normally use it to do arithmetical problems that I could do in my head if I had to.

Since my task here is not to give a proper analysis of power, I will not worry too much about how we individuate processes when we evaluate it. Control capacities can contribute to power at all sorts of levels of generality. A capacity to selectively apply a process that yields true beliefs about elms, for instance, can increase one's power of tree identification. Or it can just increase the power of that individual process, which might, for instance, increase the speed of tree-identification (if the other available processes for detecting elms are slower).

Solving any of the three metacognitive problems I described in §VI.2 can increase the power of one's capacities. Selective application includes avoiding using processes when they are likely to yield falsehoods, but also initiating them when they are likely to yield significant truths—which enhances the power of one's belief-forming capacities. Capacities for conflict resolution also extend the range of our belief-forming capacities, since they allow us to form true beliefs in cases of conflict. Much of control is directed at appropriate resource management, which is only rarely necessary for reliably formed belief, but is quite important for the efficient use of one's capacities. Resource management is particularly important for power, because whether resource-intensive processes can terminate in true beliefs often depends on effective management. How one allots study time, for instance, can substantially influence whether one can cover all the material one needs to learn.

Portability

In general, a cognitive process is portable inasmuch as it can occur or operate across different environments; a portable capacity is part of the agent's tool kit that she carries around with her. My capacity to remember the tune of "When the Saints Go Marchin' In" is highly portable, since I can perform that operation in just about any situation. Selective googling, the capacity to perform web searches and accurately judge the trustworthiness of the results, is not, since it requires an internet connection. (It is, however, very powerful.)

The term "portability" is due to Andy Clark (1997). Processes that are realized entirely within the brain tend to be more portable than those that involve external scaffolding. Clark speculates that traditional cognitive science may have assumed that computation must occur in the brain in part because of a deeper assumption that mental activity can only be realized in portable mechanisms (215-

6). However, it is important to note that portability of the raw materials does not entail portability of the process. Retrieval from LTM sometimes requires specific cues. (For instance, there are many pieces of music that I can remember in their entirety, but only if I run them through my head from the beginning; the later parts are only accessible given the cues of the earlier parts.) If retrieval of an item depends on an external cue, then even if the retrieval is entirely realized in the brain, it's not very portable; it can only occur in environments containing the appropriate cues. In contrast, logical or mathematical reasoning that uses a stylus and writing surface as an external working memory is more portable than one might think, because the external working memory can be realized by many different common objects. Blackboards are preferred, but there are always napkins. I have heard that Poincaré was struck with inspiration on a bus trip, and at each stop jumped off to write equations on the side of the bus.

For epistemological purposes, we should say that a belief-forming process is portable inasmuch as it generates true beliefs in a wide range of environments or situations. A BFP can thus fail to be portable either by not operating in a very wide range of environments or by yielding falsehoods in all but a narrow range of environments. The most natural thing to say is that the portability of a BFP is a matter of the range of environments over which it is reliable. That is the notion of portability that I will work with informally. Giving it a rigorous definition is a bit tricky, and too complicated to be worth entering into here.

Goldman (1986) identifies reliability, power, and speed as the chief BFP desiderata. However, we can see speed as valuable because it is an aspect of portability. Often, the most important constraints on belief-formation in a given situation are temporal, and processes will fail to apply in that situation if they are too slow.

Much of what I said about regarding power is also true of portability. We can understand it as a comparative notion; P_1 is more portable than P_2 iff the environments over which P_2 is reliable are a proper subset of those in which P_1 is.

Portability is strictly speaking only valuable when the process yields significant true beliefs over the additional range of environments. And we need not worry too much about individuating the processes to which appraisals of portability are applied.

Control capacities contribute to portability in much the same way that they contribute to power. Conflict resolution permits belief-formation in environments with conflicting stimuli. Positive selective application permits agents to initiate processes when they would be reliable but would not automatically be triggered. Some environments tax resources more than others. Most obviously, the environment can impose severe temporal constraints on belief formation. But it can also impose, for instance, constraints on working memory. One cannot typically realize complex logical reasoning when driving (even when pens and napkins are available), because driving successfully takes up too much attention for serious deduction to be possible. In many cases like these, successful resource management can allow belief-formation to operate in environments where it otherwise could not.

Significance

Since power and portability are each only valuable when they give rise to significant true beliefs, what we've already said covers this aspect of successful cognizing. But it may be worth highlighting separately how control capacities can contribute specifically to the formation of significant beliefs. Well-regulated inquiry is much more likely to yield answers to the important questions.

On any given subject today, there is far more work published than is feasible to read. Successful researchers need to be able to determine what is worth reading and what isn't. More precisely, they need to be able to determine how much effort to devote to any piece of literature, which can range from none to a quick skim to a careful read taking notes. By successfully navigating the literature, you can avoid forming false beliefs, but more importantly you avoid spending time on material that ultimately will not further your academic interests. Thus the value of this skill is largely its contribution to significance; it may not increase the reliability or power of your beliefs very much, but it increases the probability of having beliefs that are significant to you.

Of course, an excessive diligence in filtering out unimportant information can be deleterious. One can never quite know what will turn out to be important, particularly when what one needs is a novel solution to a problem. The discovery of penicillin started with a chance observation of a dirty Petri dish; the theory that benzene has a hexagonal structure was inspired by a dream. More interestingly (and less well known), much of Newton's mechanics can be traced back to his study of alchemy (see Peterson 1999: 420-1). Controlling for significance involves striking a mean between wasting resources acquiring useless information and wasting resources ignoring what would be useful.

Desiderata of control capacities

BFP-portability is a special case of a more general property of cognitive processes—the range of environments over which they can operate. Likewise, we can generalize power to other processes; the power of a process is (roughly) the range of appropriate outputs that it can yield. A reflex arc is quite weak, because it yields only a narrow range of motor outputs; but human motor planning can be very powerful.

Control capacities can be powerful and portable too. (They are already "reliable" in the sense that effective control capacities need to make appropriate changes to the object level most of the time.) The power of a control capacity is the range of object processes to the success of which it contributes. A general capacity for finding coherence among putative beliefs is more powerful, for instance, than a capacity just for resolving conflicts between vision and background knowledge, which is more powerful than a capacity just for determining whether I should believe the clerk at the video rental when I clearly remember having already returned the movie.

A control capacity is portable inasmuch as it can be applied in a wide range of environments. Consciously evaluating the reliability of the evidence for a proposition is powerful but not portable. You cannot consciously evaluate the reliability of your ostensible memory that bears can't climb trees when you're trying to escape an angry bear. On the other hand, judging the veracity of a statement like your companion's claim that bears can't climb trees by how familiar it seems is highly portable (being extremely fast and using minimal resources) although also not particularly reliable when not supplemented by other indications of reliability. Familiarity plays an important role in judging reliability (see Koriat 1994) perhaps just because it sometimes helps and it is nearly always available.

Just as power and portability are desiderata of BFPs, they are desiderata of control capacities. Power and portability of control capacities will (as we will see below) prove very important in understanding high-level intellectual virtues.

A control capacity can be particularly valuable by virtue of being powerful and portable. Suppose S has a knack for avoiding overly hasty belief-formation. S does not necessarily always deliberate carefully about his beliefs, but he only makes snap decisions when the situation requires it. More precisely, S has a knack for avoiding hasty belief-formation except for when the cost in significant truths is too great. (And of course he may not realize that he can do this, and his capacity may not be based in his propositional knowledge.) This sort of care in belief-formation—which is a crucial aspect of the virtue of conscientiousness—may not increase the reliability of S's belief-formation very much. It's not as if without this trait, S lacks knowledge, and with it, he has it. But if he generally manifests this trait, if it applies in most situations and to most belief-forming processes, then its value comes from the breadth of its efficacy rather than its influence on any particular case of belief-formation.

Global and local goals

Much of effective control consists in balancing different aspects of our cognitive goals. Suppose, for instance, that to solve problem P subject S has two available methods. M_1 is time-consuming and difficult, but almost certain to yield the right answer, while M_2 is a quick and dirty heuristic, requiring fewer resources but also less reliable. Effectively solving the problem in different situations depends on being able to apply the method that is most appropriate for the situation. If M_1 is initiated when there isn't time or resources for it to complete, then S might be left with no trustworthy belief or (if early termination goes bad) an untrustworthy one. But if M_2 is initiated when M_1 could be, then S is taking an unnecessary risk of acquiring a false belief by using a less-reliable process.

We should note that it can be one thing for a control capacity to allow one to attain a local goal, by allowing the agent to succeed at one aspect of the cognitive goal; and it can be another thing for a control capacity to contribute to one's overall goal of attaining significant true beliefs. This is most obvious when the local goal is the attainment of knowledge—in which case the agent has earned some credit—but the cost of the knowledge is a failure to attain some more general goals. Call a control capacity that leads to this sort of result with some regularity a *pernicious virtue*.

It's quite plausible that the need to preserve our confidence in our own abilities leads us to tend to overestimate our own chances of success. So suppose Frank has a capacity for correcting for this and other biases in the appraisal of his own cognitive limitations. In fact, suppose this is an aspect of certain control capacities that make his appraisals of his own biases and other limitations very accurate. However, given his personality, this knowledge of his own biases and limitations saps his confidence in his own abilities, making him unwilling to engage in inquiry and overly cautious in forming beliefs. His capacity for recognizing his own limitations gives him self-knowledge, but the beliefs it generates significantly reduce the power of his other belief-forming capacities.

Thus it is a pernicious virtue, yielding knowledge, but being otherwise deleterious.

Since humans do most of their cognizing in communities, we should also count the furtherance of the cognitive goals of others, or of the community as a whole, as a global goal of human cognition. Much of the value of originality, for instance, arises from how it furthers the knowledge of the whole community (Zagzebski 1996: 182-3). So we can also have pernicious virtues that yield knowledge for the individual agent but are deleterious to others around him. Suppose Karl Rove[1] has control capacities that allow him to determine reliably what lie would most effectively manipulate the beliefs of others to attain the result he desires. It can easily be verified that such a capacity would allow Rove to know that such-and-such lie will attain such-and-such result; by the account in §VII.1, then, Rove's capacity for coming up with good lies is a virtue. But lying is highly deleterious to the cognitive capacities of those who are lied to. Thus this is a pernicious virtue; it yields states that are creditworthy in a narrow and purely epistemic sense, but that are bad and blameworthy when we look at the big picture. The blameworthiness in this sort of case is to a certain extent moral; but inasmuch as the moral blame arises from depriving others of an epistemic good, the trait deserves epistemic condemnation as well.

Despite being epistemic virtues, pernicious virtues do not look all that virtuous. It might be better to call control capacities that generate knowledge "belief-forming skills" or something like that, and reserve the term virtue for traits that more closely resemble moral virtues. However, all of what I am calling "virtues" here have the same basic structure (as we'll see shortly). Moreover, it is standard practice to call knowledge-generating character traits "virtues". Thus I will continue with the terminology I have been using despite its peculiarity.

[1] Any resemblance to actual persons, living or dead, is entirely coincidental.

2. The general structure of intellectual virtue

I am now in a position to generalize my earlier definition of virtue. In §VII.1, I gave a refined definition that takes account of easy management cases—reliable clairvoyants and the like. Those refinements will also be part of our more general definition, but for ease of exposition I will suppress them for now.

So let us say that:

> S has an intellectual virtue iff she has a stable capacity to exert control over her cognitive processes in a way that allows her to form significant true beliefs and avoid forming false ones
> (equivalently: in a way that allows her to attain her cognitive goals).

Thus, a control capacity can be a virtue not just by making beliefs more reliable, but also by helping its possessor achieve some other goal of hers. Virtues can increase the power or portability of BFPs, the proportion of significant beliefs, and so on and so forth. The definition of epistemic virtue given in §VII.1 is of course the special case where the virtue is specifically a capacity to control the reliability with which a belief is produced.

When a particular process desideratum arises from an intellectual virtue, it is attributable to the subject. And, of course, control need not be exerted at any time. An agent can be in control if she would be able to correct her belief-formation in order to achieve her goals if she had to; she need not exert any special effort when she does not have to. One need not micromanage to be in control. Finally, one should note that control need not be exerted directly or by force of will. For instance, a habit of taking walks or baths at the right time can increase the power of one's problem-solving capacities.

3. Applications

Let's look at a few examples that will illustrate how this definition captures the traits we identified as virtues in ch. III. Consider what Morton (2004) calls the H and C virtues. When beliefs $B_1, \ldots, B_n$ entail p, H-virtues help us determine whether to reject $B_1 \wedge \ldots \wedge B_n$, accept p, or suspend judgment. Reasoning can branch out in any number of directions, most of which are fruitless or (though possibly interesting) irrelevant to present concerns. C-virtues help us through the maze by letting us determine which routes to trace out and which to ignore.

H-virtues can be seen as power- and reliability- encouraging control capacities. We may assume that the agent has belief-revision processes for rejecting premises or accepting conclusions; the H-virtues are capacities for controlling which of these should be engaged in at the time. They enhance reliability inasmuch as misidentifying *reductios* as interesting consequences, and vice versa, is not truth-conducive. They enhance power inasmuch as they permit beliefs to be formed at all on these grounds. Note that a situation where $B_1, \ldots, B_n$ appear to be true, p appears to follow from them, p appears to be false, and the agent is unwilling to disregard the laws of logic[2] is an informational conflict. H-virtues are involved in conflict resolution as well as selective application.

C-virtues, on the other hand, appear to be capacities for selective application and resource management. They are capacities for avoiding wasting resources on blind alleys and tangents. To be able to avoid blind alleys, of course, selective application is necessary—agents with C-virtues must be able to identify what routes are likely to yield true beliefs and which are not. C-virtues primarily enhance power and portability. They also increase the likelihood of acquiring significant true beliefs; they help us reason to the answers we wish to have, rather than to true but trivial conclusions.

[2] In case it's possible for the agent to hold $\{B_1, \ldots, B_n, B\neg p\}$ despite their inconsistency.

Aristotelian virtues

Note that our discussion in the last section was about H- and C-virtues, rather than the H-virtue and the C-virtue. Part of the reason for this is that a number of character traits can contribute to these control capacities: "conservatism, stubbornness, and doggedness" describe responses to H-type problems; "caution, foresight, stubbornness, and courage" to C-type (Morton 2004: 484-5). Perhaps more importantly, one can have local or context-dependent H- or C-virtues. One might have H-virtues leading to a refined conception of justice, but blindly accept the most ludicrous metaphysical conclusions.[3] A physicist might be very good at identifying fecund routes for reasoning regarding physics. But when she turns to philosophy of science, she might be incapable of recognizing important presuppositions of her beliefs that need to be considered; and when she plays chess, she might waste her efforts evaluating hopeless strategies and impossible contingencies. Or one might have a C-virtue that applies very generally to reasoning on different subject matters, but be easily flustered under pressure, and lose that capacity altogether. By our definition, there is no barrier to a virtue's being localized or context-dependent in these sorts of ways. (As we saw in §V.1, if we don't allow localized virtues, we lose the link between virtue and knowledge.)

Nonetheless, there is a particular epistemic value to having traits of cognitive character that resemble Aristotelian virtues—i.e., that are applicable in a very wide range of situations and to a wide range of problems. (For the reasons discussed in §V.1, I'll disregard the motivational component of Aristotelian virtues.) Such characteristics are valuable specifically because of their own power and portability (in addition to their contributions to the power and portability of object-level processes). In the next two sections, we'll look at two examples—originality and humility—to see how we can account for high-level virtues on the theory I am advancing.

[3] No, there isn't anyone in particular I have in mind here.

Originality

Originality presumably involves certain fairly basic attributes of a cognitive system that help generate novel hypotheses. For instance, Carson et al. (2003) found a link between difficulties with latent inhibition—the capacity to screen from consciousness stimuli already determined to be irrelevant—and creative thinking. Original thinkers might also engage in activities that encourage the development of novel ideas. Brainstorming is one; leisurely, unpressured thinking is another.

However, originality involves more than just a capacity to generate novel ideas. Difficulties with latent inhibition are also linked to psychosis. One must also be able to recognize when novel ideas can serve as the basis for solutions to problems—when they are worth developing or believing, and when they are not. Most ways of being original are wrong, and most novel ideas are mere distractions. The thinker with lots of ideas must be attracted enough to novelty to develop novel solutions, but not so much that she adopts crazy beliefs.

Thus the original thinker is distinguished from the distractible or reckless thinker by the capacity to *control* the application of novel ideas to problem solving. The original thinker is prone to having novel ideas, and may even be able to control this disposition to a certain extent, fostering it when useful and suppressing it when distractions are likely to be costly. But the original thinker is also capable of developing those ideas, recognizing which ones are potentially valuable and turning them into working, believable solutions when necessary. Thus the virtue of originality can be seen as a power-enhancing control capacity for managing the production of beliefs based on novel ideas.

Humility

Let us now turn to humility. Roberts and Wood (2003) argue (compellingly) that intellectual humility is a dispositional lack of concern with the status that goes with intellectual achievements or with dominating the thinking of others, and a disposition not to claim unwarranted entitlements on the basis of one's intellectual excellence. Most importantly, the intellectually humble person prefers knowledge and other epistemic goods to status, entitlement, and domination, and will not sacrifice her epistemic goals for recognition or influence. The chief vices opposed to humility are vanity, arrogance, and pretension.

Humility is a moral as well as intellectual virtue, and much of its value consists in its contribution to a good life. Roberts and Wood propose, however, that humility furthers our cognitive ends as well. Humans engage in intellectual endeavours in communities. Not only do we acquire much of our knowledge from testimony, but others provide much of the critical analysis that our beliefs so often need to be trustworthy. The humble cognizer prefers attaining epistemic goods to impressing others; she is not unwilling to lose face by admitting error or acknowledging that others are right. She does not dismiss testimony or criticism because it comes from those whom she perceives as her intellectual inferiors. This habit is advantageous mainly because professional status and renown are poor indicators of the reliability of any given belief. The humble cognizer does not "bar the views of others from consideration" (273), and thus is fully open to the contributions that others can make to the refinement and revision of her beliefs.[4]

Status and dominance are basic human goals in some sense (and I will not comment here on whether or when they ought to be pursued). Intellectual humility can be seen as a stable character trait that controls belief-formation to prevent it from being biased by desires for status and dominance. It is powerful and portable, since it applies generally to the agent's BFPs and is stable across scenarios. It only influences one aspect of belief-formation, and thus must be conjoined with other control capacities in order to yield knowledge. Humility

[4] Intellectual humility, like many other virtues (especially those that overlap with moral virtues), makes special contributions to the success of inquiry conducted in groups. Thus it has further advantages that I do not discuss above.

guards against certain biases but does not guarantee reliability. It is epistemically valuable not because it dependably generates knowledge, but because it dependably encourages epistemic goods across belief-formation.

The humble person can be in control of these biases even if this doesn't require causal intercession into belief-formation, as we saw in §VII.1. It's not as if to exhibit humility, one must (say) feel the temptation to dismiss a seeming inferior, but withstand the urge. One can be in control as long as it is the case that were desires for status and dominance to well up, one would prevent them from deleteriously influencing one's beliefs. As long as the capacity for interceding to prevent bias is there, it may be necessary only very rarely, or even only counterfactually.

Why regard humility as a control capacity and not just insensitivity to considerations of status and dominance? To begin with, the latter option diverges substantially from theories of moral virtue. Aristotle distinguishes between "natural virtues"—unlearned capacities to be correctly motivated or make virtuous decisions—and virtues *per se*, which must be developed through appropriate training and practice. Likewise, we have here a distinction between an agent who can control for biases in belief-formation and an agent who simply does not feel any desire for intellectual status or dominance, or whose belief-forming processes are naturally disconnected from such desires.

Suppose, for instance, we designed a thinking robot and explicitly program it so that it has no desires for status or intellectual domination. It considers all opinions without considering the perceived status of their source, because it is incapable of bringing considerations of perceived status to bear on its inquiries. Intuitively, "humble" is entirely the wrong word to describe the robot. It is better to say that it conducts inquiry dispassionately. If it does not have the capacity for those desires to mislead it (and thus lacks the capacity to prevent them from doing so), then it is not right to call it humble. This illustrates how humility is something different from mere insensitivity. Taking intellectual virtues to be control capacities allows us to give a proper account of how they differ from mere fortuitous natural traits.

Intellectual success and intellectual virtue

As we saw in §III.2, teleological accounts of virtue must explain why our appraisals of virtue sometimes differ from our appraisals of the frequency and breadth of true beliefs. The divergence between virtue and consequences is most obvious with past agents. Our present understanding of certain areas—particularly the physical and deductive sciences—vastly outstrips our forebears'. Educated persons today are more likely to be right on a wider range of questions than even the great minds of the past. People like Newton or Aristotle are regarded as exemplars of intellectual virtue (at least with regards to the sciences), despite the fact that the reliability and power of their faculties are unexceptional compared to educated people today, and in some cases clearly deficient. To consider another comparison, Newton and Einstein are not far apart in degree of virtue, but the latter's beliefs were far more likely to be true than the former (see Riggs 2003b: 210-3).

What makes it possible to explain our appraisals here is the fact that a virtuous trait can be valuable in part because of its own power and portability, even though it does not typically have enough effect on belief-formation to make one's beliefs into knowledge. We saw an example of this above with humility. Just because a belief arises in part from humility—in that, say, a nonhumble person would have dismissed the person who gave the testimony—does not make it knowledge. The testifier must also at least be reliable. Newton's conscientiousness and originality were certainly conducive both to reliability and to power, as is illustrated by the greater reliability and power of his views over most of his contemporaries. But this is not sufficient to give him as much knowledge as a later scientist could acquire, given the limited background knowledge he had to draw on, the lack of suitable instruments, etc.

On the other hand, a well-educated but otherwise dull scientist today has a wide range of reliable processes that she can use to get knowledge. But she will lack the sort of general control capacities that are so praiseworthy in Aristotle and Newton. This will impugn the reliability and power of her belief-formation. Nonetheless, the relative advantage of her low-level mechanisms can easily outweigh the relative advantage of the high-level virtues of great thinkers of the past.

This account captures our evaluations in this case quite adequately. Note that we are ambivalent in our appraisals of past thinkers. We hold up Aristotle's conscientiousness in logic as a trait to emulate, but we don't emulate his use of syllogistic reasoning, which is weak and not entirely reliable. Likewise, Newton is a scientific genius, but not a scientific authority; he is to be emulated but not trusted without corroboration. These commonsense evaluations match their status on the position being advanced here, of having some virtues that most of us today lack, but lacking some virtues we can readily acquire.

4. Towards a more complete account of virtues

The examples and arguments adduced in the last section should establish that the account of virtue I am urging can account for both high-level and low-level virtues. This is a significant advantage of the account; as we saw in ch. IV, extant theories of virtue do not do a very good job of unifying the two levels of virtue. My analysis suggests, moreover, many intellectual virtues have been overlooked so far. There should be many localized virtues, applying only in a certain range of contexts. There should also be a plethora of resource-management virtues, or skills we have for keeping problem solving tractable.

These hitherto-unnoticed virtues play an important role in our understanding of our own and others' cognition. For instance, suppose Quincy exhibits intellectual humility in calm situations when he feels confident. But when he is stressed or pressured, or after what he takes to be a direct attack on his person or capacities, he overcompensates with arrogance. Quincy lacks the virtue of humility as this is traditionally understood, since his humility is not a robust character trait. He has a localized virtue of humility—a control capacity that functions like humility but that applies in a narrower range of circumstances. Being less portable, Quincy's trait is less valuable than a robust humility, but it can nonetheless be important to identify it and the value it has. It means, for instance, that Quincy can be just valuable for group problem solving as a truly humble person would, as long as he is not pressured or offended.

Quincy's humility is much like a capacity to acquire knowledge in a certain context. It's important to identify such capacities because we can trust the person with that localized skill when she is in the right context. Likewise, we know that the environment has to be correctly set up for Quincy to be humble, but when it is, his presence can be very valuable. Quincy's humility might even be something thin-skinned persons might aspire to, if they believe that they cannot expect themselves to remain humble under pressure or attack but wish to attain the best character traits they can.

Standard theories of virtue miss traits like these, since they are neither robust nor produce knowledge. Nonetheless, they are an important part of how we evaluate cognitive character, and thus cannot be dismissed.

I have only given an account of the abstract structure of intellectual virtues here. A great deal more work needs to be done on categorizing and understanding the virtues, both the traditionally recognized ones and more localized or less obvious virtues. Most importantly, we need a better sense of just what roles these many virtues play in epistemic evaluation and how different process desiderata are interrelated (including their relative importance). But these are tasks for future research.

5. Conclusion

I will conclude by summarizing the view of intellectual virtue for which I have argued. Intellectual virtues are capacities for controlling one's own cognitive processes so that one can attain the epistemic goal of significant true belief. This account of virtue has two chief advantages.

First, on this account there is a clear sense in which virtuously held belief proceeds from an agent's own cognitive character and the agent can be credited for it. Suppose S's belief B has epistemic desideratum D and S has a capacity to control the processes that generated B so that they tend to yield beliefs with desideratum D. Then we can say that S has an ability to generate beliefs with desideratum D (using those processes, in conditions that normally obtain) and thus that his belief's having that desideratum is his own doing. As I argued in §VII.1, reliable mechanisms that do not yield knowledge fail to do so because they are not under effective control. When, e.g., a reliable but uncontrolled faculty does yield true belief, the belief's being true is not due to the subject's own efforts and powers and thus fails to be knowledge.

At the beginning of ch. IV, I endorsed what Alston calls the "imperialist pretensions" of virtue epistemology—the claim that intellectual virtue occupies a central place in the understanding of the evaluation of beliefs. The second advantage of my account is that it gives victory to the imperialists. The chief barrier to taking virtue to be central to epistemology is the apparent distinction between epistemic or low-level virtues (which are necessary for knowledge) and high-level virtues (which are only peripherally involved in knowledge). My analysis allows us to give a unified account of both types of virtue, as well as other traits of epistemic activity that virtue epistemologists have not yet studied. Control capacities can contribute to the end of significant true belief by helping one's processes generate reliable beliefs (which I argued in §VII.1 is necessary for knowledge), or by rendering one's processes more powerful, portable, or likely to generate significant beliefs. Control capacities can themselves have extra value by being powerful or portable, as Aristotelian virtues are. Because metacognition can contribute to epistemic value in many ways, many different appraisals of subjects and beliefs can be understood as evaluations of metacognitive capacities.

It is thus to be hoped that the theory presented in this dissertation will help provide a rigorous foundation for the study of intellectual virtue, and help establish the importance of intellectual virtue to epistemology.

WORKS CITED

Adler, J. E. (2002) *Belief's own ethics.* Cambridge, MA: MIT.

Adler, J. E., and M. Levin (2002) "Is the generality problem too general?" *Philosophy and Phenomenological Research* 65(1): 87-97.

Alston, W. P. (1993) "Epistemic desiderata". *Philosophy and Phenomenological Research* 53(3): 527-51.

———. (1995) "How to think about reliability". *Philosophical Topics* 23(1): 1-29.

———. (1996) *A realist conception of truth.* Ithaca, NY: Cornell.

———. (2005) *Beyond "justification": Dimensions of epistemic evaluation.* Ithaca, NY: Cornell.

Annas, J. (2003) "The structure of virtue". In DePaul and L. Zagzebski, eds., 15-33.

Anthony, C. K. (1988) *A guide to the I Ching.* Anthony Publishing Co.

Aristotle (350BC/1980) *The Nicomachean ethics.* David Ross, trans. Oxford: Oxford.

Axtell, G., ed. (2000) *Knowledge, belief, and character: Readings in virtue epistemology.* Lanham, MD: Rowman & Littlefield

———. (2006) "Blind man's bluff: The basic belief apologetic as anti-skeptical stratagem". *Philosophical Studies* 130: 131-52.

Bach, K. (1985) "A rationale for reliabilism". *Monist* 68: 248-63.

Beebe, J. R. (2004) "The generality problem, statistical relevance and the tri-level hypothesis". *Noûs* 38(1): 177-95.

Benjamin, A., and R. Bjork (1996) "Retrieval fluency as a metacognitive index". In L. M. Reder, ed. *Implicit memory and metacognition.* Mahwah, NJ: Erlbaum, 309-38.

Bernecker, S. (200x) "Agent reliabilism and the problem of clairvoyance". *Philosophy and Phenomenological Research*, forthcoming.

BonJour, L. (1980) "Externalist theories of empirical knowledge". *Midwest Studies in Philosophy* 5: 53-73.

———. [with E. Sosa] (2003) *Epistemic justification: Internalism vs. externalism, foundations vs. virtues.* Oxford: Blackwell.

Brooks, R. A. (2002) *Flesh and machines: How robots will change us.* New York: Pantheon.

Brown, J. R. (2001) *Who rules in science: An opinionated guide to the wars.* Cambridge, MA: Harvard.

Byron, M. (1998) "Satisficing and optimality". *Ethics* 109(1): 67-93.

Carson, S. H., J. B. Peterson, and D. M. Higgins (2003) "Decreased latent inhibition is associated with increased creative achievement in high-functioning individuals". *Journal of Personality and Social Psychology* 85(3): 499-506.

Cary, M., and L. M. Reder (2002) "Metacognition in strategy selection". In Chambres, Izaute, & Marescaux, eds., 63-77.

Ceci, S. J. (1993) "Contextual trends in intellectual development". *Developmental Review* 13: 403-35.

———. (1996a) "General intelligence and life success: An introduction to the special theme". *Psychology, Public Policy, and Law* 2(3/4): 403-17.

———. (1996b) *On intelligence: A bioecological treatise on intellectual development.* Cambridge, MA: Harvard.

Chambres, P., M. Izaute, & P.-J. Marescaux, eds. (2002) *Metacognition: Process, function, and use.* Dordrecht: Kluwer.

Clark, A. (1997) *Being there: Putting brain, body, world together again.* Cambridge, MA: MIT.

———. (2003) *Natural-born cyborgs: Minds, technologies, and the future of human intelligence.* Oxford: Oxford.

Clark, A., & J. Toribio (1994) "Doing without representing?" *Synthese* 101(3): 401-31.

Code, L. (1987) *Epistemic responsibility.* Hanover, NH: Brown.

Cohen, S. (2002) "Basic knowledge and the problem of easy knowledge". *Philosophy and Phenomenological Research* 65(2): 309-29.

Comesaña, J. (2006) "A well-founded solution to the generality problem". *Philosophical Studies* 129(1): 27-47.

Conant, R. C., and W. R. Ashby (1970) "Every good regulator of a system must be a model of that system". *International Journal of Systems Science* 1(2): 89-97.

Conee, E., and R. Feldman (1998). "The generality problem for reliabilism". *Philosophical Studies* 89: 1-29.

Cornoldi, C. (1998) "The impact of metacognitive reflection on cognitive control". In G. Mazzoni and T. O. Nelson, eds. *Metacognition and cognitive neuropsychology: Monitoring and control processes.* Mahwah, NJ: Erlbaum.

Dancy, J. (1985) *An introduction to contemporary epistemology.* Malden, MA: Blackwell.

Danto, A. C. (1967) "On knowing that we know". In A. Stroll, ed. *Epistemology: New essays in the theory of knowledge.* New York: Harper, 32-53.

Darling, S., S. D. Sala, C. Gray, and C. Trivelli (1998) "Putative functions of the prefrontal cortex: Historical perspectives and new horizons". In G. Mazzoni & T. O. Nelson, eds. *Metacognition and cognitive neuropsychology: Monitoring and control processes.* Mahwah, NJ: Erlbaum, 53-95.

Davidson, D. (1986) "A coherence theory of truth and knowledge". In E. LePore, ed. *Truth and interpretation.* Malden, MA: Blackwell, 307-19.

Davies, M., and M. Coltheart (2000) "Introduction: Pathologies of belief". *Mind and Language* 15(1): 1-46.

Dawson, M. R. W. (1998) *Understanding cognitive science.* Malden, MA: Blackwell.

DePaul, M., and L. Zagzebski, eds. (2003) *Intellectual virtue: Perspectives from ethics and epistemology.* Oxford: Oxford.

Doris, J. M. (1998) "Persons, situations, and virtue ethics". *Noûs* 32(4): 504-30.

Driver, J. (2000) "Moral and epistemic virtue". In Axtell, ed., 123-34.

Fairweather, A., and L. Zagzebski, eds. (2001) *Virtue epistemology.* Oxford: Oxford.

Foley, R. (1987) *The theory of epistemic rationality.* Cambridge, MA: Harvard.

Gigerenzer, G., P. M. Todd, and the ABC Research Group (1999) *Simple heuristics that make us smart.* Oxford: Oxford.

Gilbert, D. T. (1991) "How mental systems believe". *American Psychologist* 46(2): 107-19.

Goldman, A. (1980) "The internalist conception of justification". *Midwest Studies in Philosophy* 5: 27-51.

———. (1986) *Epistemology and cognition.* Cambridge, MA: Harvard.

———. (1992) *Liaisons: Philosophy meets the cognitive and social sciences.* Cambridge, MA: MIT.

———. (1994) "Naturalistic epistemology and reliabilism". *Midwest Studies in Philosophy* 19: 301-20.

———. (2005) "Disagreement in philosophy". In H. D. Battaly and M. P. Lynch, eds. *Perspectives on the philosophy of William P. Alston.* Lanham, MD: Rowman & Littlefield.

Greco, J. (1990) "Internalism and epistemically responsible belief". *Synthese* 85: 245-77.

———. (1993) "Virtues and vices of virtue epistemology". *Canadian Journal of Philosophy* 23(3): 413-32.

———. (2000a) *Putting skeptics in their place.* Cambridge, UK: Cambridge.

———. (2000b) "Two kinds of intellectual virtue". *Philosophy and Phenomenological Research* 60(1): 179-84.

———. (2001) "Virtues and rules in epistemology". In Fairweather and Zagzebski, eds., 117-41.

———. (2003a) "Further thoughts on agent reliabilism". *Philosophy and Phenomenological Research* 64(2): 466-88.

———. (2003b) "Knowledge as credit for true belief". In DePaul and L. Zagzebski, eds., 112-34.

———, ed. (2004) *Ernest Sosa and his critics.* Malden, MA: Blackwell.

———. (2004) "How to preserve your virtue while losing your perspective". In J. Greco, ed., 96-105.

Grimm, S. R. (2001) "Ernest Sosa, knowledge, and understanding". *Philosophical Studies* 106: 171-91.

———. (200x) "Epistemic goals and epistemic values". *Philosophy and Phenomenological Research*, forthcoming.

Grush, R. (2003) "In defense of some 'Cartesian' assumptions concerning the brain and its operation". *Biology and Philosophy* 18(1): 53-93.

Harman, G. (1999) "Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error". *Proceedings of the Aristotelian Society* 99: 315-31.

Hasher, L. (1977) "Frequency and the conference of referential validity". *Journal of Verbal Learning and Verbal Behaviour* 16: 107-12.

Hawthorne, J. (2004) *Knowledge and lotteries.* Oxford: Oxford.

Hegel, G. W. F. (1805) *Philosophy of making stuff up.* Jayson Blair, trans. Dordrecht: Kluwer.

Hookway, C. (2003) "How to be a virtue epistemologist". In DePaul and L. Zagzebski, eds., 183-202.

Hunter, J. E., and F. L. Schmidt (1996) "Intelligence and job performance: Economic and social implications." *Psychology, Public Policy, and Law* 2(3/4): 447-72.

Hurley, S. L. (1998) *Consciousness in action.* Cambridge, MA: Harvard.

Kitcher, P. (1992) "The naturalists return". *Philosophical Review* 101(1): 53-114.

Koriat, A. (1994) "Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing". In Metcalfe & Shimamura, eds., 116-35.

Kuhn, T. S. (1970) *The structure of scientific revolutions.* Chicago: Chicago.

Lackey, J. (200x) "Why we don't deserve credit for everything we know". *Synthese*, forthcoming.

Lehrer, K. (1974) *Knowledge.* Oxford: Oxford.

———. (1990) *Metamind.* Oxford: Oxford.

Lepock, C. (2006) "Adaptability and perspective". *Philosophical Studies* 129(2): 377-91.

Metcalfe, J., & A. P. Shimamura, eds. (1994) *Metacognition: Knowing about knowing.* Cambridge, MA: MIT.

Millikan, R. G. (1993) *White Queen psychology and other essays for Alice.* Cambridge, Mass.: MIT.

Mitchell, K. J., and M. K. Johnson (2000) "Source monitoring: Attributing mental experiences". In E. Tulving & F. M. Craik, eds. *The Oxford handbook of memory.* New York: Oxford, 179-95.

Montmarquet, J. A. (1993) *Epistemic virtue and doxastic responsibility.* Lanham, MD: Rowman & Littlefield.

———. (2000) "An 'internalist' conception of epistemic virtue". In G. Axtell, ed., 135-47.

Morton, A. (2003) *The importance of being understood.* New York: Routledge.

———. (2004) "Epistemic virtues, metavirtues, and computational complexity". *Noûs* 38(3): 481-502.

Nelson, T. O., and L. Narens (1990) "Metamemory: A theoretical framework and new findings". *Psychology of Learning and Motivation* 26: 125-73.

Nichols, S., S. Stich, and J. M. Weinberg (2003) "Metaskepticism: Meditations in ethno-epistemology". In S. Luper, ed. *The skeptics.* Aldershot, UK: Ashgate, 227-47.

O'Regan, J. K., and A. Noë (2001) "A sensorimotor account of vision and visual consciousness". *Behavioral and Brain Sciences* 24: 939-1031.

Paris, S. G. (2002) "When is metacognition helpful, debilitating, or benign?" In Chambres, Izaute, & Marescaux, eds., 105-20.

Peterson, J. B. (1999) *Maps of meaning: The architecture of belief.* New York: Routledge.

Plantinga, A. (1993a) *Warrant and proper function.* New York: Oxford.

———. (1993b) *Warrant: The current debate.* New York: Oxford.

———. (1993c) "Why we need proper function". *Noûs* 27(1): 66-82.

Plato (ca. 400BC/1976). *Meno.* G. M. Grube, trans. Indianapolis: Hackett.

Putnam, H. (1983) "Why reason can't be naturalized". In *Realism and reason.* Cambridge, UK: Cambridge, 229-47.

Riggs, W. D. (2002) "Reliability and the value of knowledge". *Philosophy and Phenomenological Research* 64(1): 79-96.

———. (2003a) "Balancing our epistemic goals". *Noûs* 37(2): 342-52.

———. (2003b) "Understanding 'virtue' and the virtue of understanding". In DePaul and Zagzebski, eds., 203-26.

Roberts, R. C., and W. J. Wood (2003) "Humility and epistemic goods". In DePaul & Zagzebski, eds., 257-79.

Russell, B. (1912) *The problems of philosophy.* Oxford: Oxford.

———. (1946) *A history of Western philosophy.* London: Unwin.

Schunn, C. D., and L. M. Reder (1998) "Strategy adaptivity and individual differences". *Psychology of Learning and Motivation* 38: 115-54.

Simon, H. A. (1982) *Models of bounded rationality vol. 2.* Cambridge, MA: MIT.

Sosa, E. (1980) "The raft and the pyramid: Coherence versus foundations in the theory of knowledge". *Midwest Studies in Philosophy* 5: 3-25.

———. (1991) *Knowledge in perspective.* Cambridge, UK: Cambridge.

———. (1993) "Proper functionalism and virtue epistemology". *Noûs* 27(1): 51-65.

———. (1995) "Perspectives in virtue epistemology: A response to Dancy and BonJour". *Philosophical Studies* 78(3): 221-35.

———. (1997) "How to resolve the Pyrrhonian problematic". *Philosophical Studies* 85: 229-49.

———. (2001) "For the love of truth?" In Fairweather and L. Zagzebski, eds., 49-62.

———. [with L. BonJour] (2003a) *Epistemic justification: Internalism vs. externalism, foundations vs. virtues.* Oxford: Blackwell.

———. (2003b) "Epistemology and primitive truth". In M. P. Lynch, ed. *The nature of truth: Classic and contemporary perspectives.* Cambridge, MA: MIT, 641-62.

———. (2003c) "The place of truth in epistemology". In DePaul and L. Zagzebski, eds., 155-79.

———. (2004) "Replies". In Greco, ed., 275-325.

Sreenivasan, G. (2002) "Errors about errors: Virtue theory and trait attribution". *Mind* 111: 47-68

Steup, M. (2004) "Internalist reliabilism". *Philosophical Issues* 14: 403-25.

Swanton, C. (1993) "Satisficing and virtue". *Journal of Philosophy* 90(1): 33-48.

Ugel, E. "The lottery's next big loser: Illinois". *New York Times* 28 Jan 2007: §4 p. 17.

van Cleve, J. (1984) "Reliability, justification, and the problem of induction". *Midwest Studies in Philosophy* 9: 555-67.

Vogel, J. (2000) "Reliabilism leveled". *Journal of Philosophy* 97: 602-23.

Wallis, C. (1994) "Truth-ratios, process, task, and knowledge". *Synthese* 98: 243-69.

Williamson, T. (2000) *Knowledge and its limits.* Oxford: Oxford.

Wilson, R. A. (2003) "Individualism". In S. P. Stich and T. A. Warfield, eds. *The Blackwell guide to philosophy of mind.* Oxford: Blackwell, 256-87.

Zagzebski, L. (1993) "Religious knowledge and the virtues of the mind". In L. Zagzebski, ed. *Rational faith.* Notre Dame: Notre Dame.

———. (1996) *Virtues of the mind.* Cambridge, UK: Cambridge.

———. (2000) "From reliabilism to virtue epistemology". In G. Axtell, ed., 113-22.

———. (2003) "Intellectual motivation and the good of truth". In DePaul and L. Zagzebski, eds., 135-54.

———. (2004) "Epistemic value monism". In Greco, ed., 190-8.

Zalabardo, J. L. (200x) "Externalism, skepticism, and the problem of easy knowledge". *Philosophical Review*, forthcoming.