✳

VIRTUE
EPISTEMOLOGY

*Essays on Epistemic Virtue
and Responsibility*

✳

Edited by

ABROL FAIRWEATHER
LINDA ZAGZEBSKI

Virtue Epistemology: Essays on Epistemic Virtue and Responsibility

Abrol Fairweather
Linda Zagzebski,
Editors

OXFORD UNIVERSITY PRESS

VIRTUE EPISTEMOLOGY

Essays on Epistemic Virtue and Responsibility

Edited by
Abrol Fairweather &
Linda Zagzebski

OXFORD
UNIVERSITY PRESS

CONTENTS

CONTRIBUTORS

Robert Audi, Charles J. Mach Distinguished Professor, University of Nebraska–Lincoln

Guy Axtell, Adjunct Professor of Philosophy, University of Nevada–Reno

Heather D. Battaly, Instructor of Philosophy, California State University, Fullerton

Simon Blackburn, Edna J. Koury Distinguished Professor of Philosophy, University of North Carolina, Chapel Hill

Abrol Fairweather, Instructor of Philosophy, University of San Francisco

Richard Foley, Professor of Philosophy and Dean of the Faculty of Arts and Sciences, New York University

Alvin I. Goldman, Regents Professor of Philosophy and Research Professor of Cognitive Science, University of Arizona

John Greco, Professor of Philosophy, Fordham University

Christopher Hookway, Professor of Philosophy, University of Sheffield

Hilary Kornblith, Professor of Philosophy, University of Vermont

Keith Lehrer, Regents Professor of Philosophy, University of Arizona

Ernest Sosa, Romeo Elton Professor of Natural Theology and Professor of Philosophy, Brown University. Distinguished Visiting Professor, Rutgers University

Linda Zagzebski, Kingfisher College Chair of the Philosophy of Religion and Ethics and Professor of Philosophy, University of Oklahoma

VIRTUE EPISTEMOLOGY

I

INTRODUCTION

Linda Zagzebski and Abrol Fairweather

1. A Short History of Virtue Epistemology

The name "virtue epistemology" has come to designate a class of recent theories that focus epistemic evaluation on properties of persons rather than properties of beliefs or propositions. The direction taken by this approach and the issues it raises are strikingly different from those that dominated American epistemology at the beginning of the last quarter of the twentieth century. At that time it was almost always taken for granted that knowledge is justified true belief, and epistemic discourse was dominated by competing analyses of the concept of justification. The demise of the definition had already been initiated when Edmund Gettier published his famous essay in 1963,[1] generating a long series of attempts to respond to his counterexamples without giving up the essence of the definition. Perhaps the best of these attempts was the defeasibility theory, which was proposed when it was noticed that in typical Gettier cases in which one has a justified true belief B that is not knowledge, B depends upon or otherwise "goes through" a false proposition. When the false proposition is corrected and added to the reasons justifying B, B is no longer justified.[2]

But the aftermath of the Gettier literature was the realization that the concept of justification itself was in trouble. Problems with understanding the nature of justification hardened into a controversy over the extent to which the conditions for justification are external or internal to the consciousness of the believer, and this dispute led some philosophers to separate the concepts of justification and knowledge, giving an internalist account of the former and an externalist account of the latter.[3] Even if such a move were successful, however, it would have meant trouble for justification since justification had been deemed important largely because it was thought to be a component of knowledge. In any case, such a move meant the end of the justified-true-belief (JTB) definition of knowledge. Epistemology became increasingly fragmented, and by the nineties the internalism/externalism dispute had reached an impasse, leading at least one major epistemologist to the conclusion that

the conflict was irresolvable because there was no single target about which competing theorists were making differing claims.[4]

Even more radical pronouncements on the demise of epistemology came from the death-of-epistemology theorists who maintained that the issues constituting professional epistemology had been dictated by the perceived need to respond to skepticism.[5] Once the presuppositions behind the skeptical challenge are given up, they argued, most of epistemology becomes pointless. Subsequent history indicates that the grip of skepticism on the philosophical imagination has weakened, but it has by no means disappeared. Still, it is worth noting that the preoccupation with justification that marked professional epistemology at that time was connected with skepticism since being justified is the state one desires in order to defend one's right to be sure.[6]

The motive to avoid skepticism was the impetus for another dispute that dominated epistemology during the last decades of the twentieth century — the dispute between foundationalism and coherentism on the nature of a rational cognitive structure. This dispute also appeared to be intractable, and by 1980 Ernest Sosa proposed in his important essay, "The Raft and the Pyramid," that the concept of intellectual virtue could be used to bypass the controversy between foundationalists and coherentists.[7] In that essay Sosa introduced the term "intellectual virtue" into the contemporary epistemological literature. What Sosa meant by an intellectual virtue was a reliable belief-forming faculty, and so virtue epistemology (VE) began as a species of reliabilism. According to reliabilist theories, what makes a true belief an instance of knowledge (or alternatively, what makes a belief justified) is that it arises out of a reliable faculty[8] or process[9] for obtaining the truth. Reliabilism is externalist in that the conditions for knowledge or justifiedness need not be accessible to the consciousness of the believer. Reliabilist forms of VE have little or no connection with virtue ethics.

The older JTB theory of knowledge was consciously normative; to be justified is to be in an evaluatively positive state. The concept of epistemic justification was modeled on moral justification, which in turn was commonly understood in terms of doing one's epistemic duty.[10] Reliabilism entered philosophical discourse as a competitor to the JTB theory and it was naturalistic since it held that normative epistemic properties are reducible to natural, non-epistemic properties. Reliabilism was therefore both a form of externalism and of naturalized epistemology.

In addition to reliabilism, the development of VE was influenced by the work of Lorraine Code and James Montmarquet.[11] What distinguished Code and Montmarquet from the reliabilists was that they both treated intellectual virtue on the classical model of virtue as a trait of character such as open-mindedness or intellectual fairness. Both stressed the importance of being a responsible and conscientious believer, and Code focused on the importance of the knowing subject in an epistemic community. Neither theory was allied with externalism or naturalism in epistemology.

Some more recent versions of VE remain forms of reliabilism, such as John Greco's agent reliabilism. In Greco's theory, an agent's true belief p has the value that converts true belief into knowledge just in case his believing p results from stable

and reliable dispositions that make up his cognitive character.[12] These dispositions are those he manifests when thinking in a way motivated by the attempt to get truth. Greco intends this definition to entail the satisfaction of conditions of subjective responsibility as well as objective reliability. It is therefore a form of reliabilism and is not modeled on virtue ethics, but it makes internal conditions for epistemic value crucial.

My (Zagzebski's) version of VE is explicitly modeled on virtue ethics.[13] Like Code and Montmarquet, I think of intellectual virtues as traits such as intellectual autonomy and courage, intellectual carefulness and fairness, and open-mindedness, but like Sosa and Greco, I regard reliability as a component of virtue. An intellectual virtue, like a moral virtue, has a motivational component as well as a component of reliable success in reaching the end (if any) of the motivational component. What makes intellectual virtues intellectual is that they (or most of them) include motive dispositions connected with the motive to get truth, and reliability is entailed by the success component of the virtue. This strategy shows how the internalist feature of responsibility and the externalist feature of epistemic success can be combined in a unified concept—indeed, a concept that has a long history in ethics. In my view, justification is not the most important concept in epistemic evaluation; a justified belief ought to be analyzed as the parallel of a right act in pure virtue ethics. The issue of whether a rational cognitive structure is foundationalist or coherentist is also a derivative matter, determined by what intellectually virtuous persons do. The evaluative component of knowledge is not justification, but what I call an "act of intellectual virtue." The theory is normative, but it can be interpreted as naturalistic in the sense in which Aristotle's ethics is naturalistic. That is, it does not reduce epistemic evaluative properties to natural properties, but what counts as a virtue, whether moral or epistemic, is intimately connected with the way human beings are constructed by nature.

As we have seen, then, "virtue epistemology" applies to theories that cut across divisions between externalists and internalists, foundationalists and coherentists, and normative vs. naturalistic epistemologies. Virtue epistemologists differ on the importance of justification, but none makes it the focus of the theory. Virtue epistemologists also differ on the importance of skepticism. Greco argues that VE, along with other forms of reliabilism, has the advantage of securing knowledge against skeptical threats, but other virtue epistemologists prefer to leave skeptical worries aside in order to pursue a program that is not dominated by these worries.[14]

The essays in this volume are responses to the ascendancy of virtue epistemology. Some authors are already known for their work in VE (Sosa, Zagzebski, Greco, Hookway), or for a theory closely associated with it (Goldman). Others are making contributions to it for the first time, and some of these are well known for their work outside VE. Two authors known outside VE (e.g., Foley, Kornblith) are addressing the normativity of epistemology from a different direction, but all the essays illustrate how the scope of normativity in epistemology has expanded in recent years. Justification is a secondary interest in virtually every essay in this volume and even the exceptions are enlightening. Battaly examines problems in the concept of justification, but in order to caution virtue epistemologists not to fall into the same

sort of problems over the concept of virtue. Audi argues that justification and knowledge can be illuminated by an investigation of the parallels between moral and intellectual virtue. The one essay focused on epistemic obligation (Kornblith) does not connect it with justification. None of the essays gives more than passing attention to skepticism. Few address the division between foundationalism and coherentism, and only Kornblith and Axtell say much about the internalism/externalism dispute. However, there is considerable discussion of concepts related to epistemic agency, including responsibility, credit, negligence, control, habit, goals, motives, rule-following, obligation, and even *akrasia*.

2. Summary of Chapters

In "Reason, Virtue, and Knowledge," Simon Blackburn explores some of the relations between a virtue approach to epistemology and a minimalist or deflationist conception of truth. To be interesting, Blackburn points out, VE must defend the priority of the concept of epistemic virtue over the concepts of justification, knowledge, or truth, in ascending order of strength, just as an interesting virtue ethics must defend the priority of moral virtue over the concepts of right act and good outcome. But a difficulty confronts the virtue theorist. If, as Blackburn believes, reliability sits firmly in the center of cognitive virtues, then the priority she needs to defend seems to be reversed. For then a trait gets to be on the list of epistemic virtues because it promotes an alignment of belief and truth. This is parallel to saying that a trait counts as a virtue because it promotes utility or wards off loss, where utility and loss are independently understood. In order to defend its priority, it appears that virtue epistemology will need a robust or thick conception of truth, where truth is valuable and intellectual traits are classified as virtues insofar as they lead to it.

Blackburn argues, however, that this is not the case, for a minimalist or deflationist theory of truth can, perhaps surprisingly, deliver the requisite sense of the value of truth in a moderately strong VE. The virtues are handmaidens to truth, but Blackburn tentatively concludes that we need not suppose that the relationship between virtue and truth undermines moderately strong epistemic virtue theory provided that the theory includes both minimalism about truth and a version of the "use theory" of meaning in which use is primarily identified by virtuous verification or assertibility conditions.

Alvin Goldman's essay, "The Unity of the Epistemic Virtues," explores the proposition that the various epistemic virtues are "unified" in the sense that they are all variations on, or permutations of, a single theme: achieving a high degree of success on questions of interest. All epistemic virtues attain their status as virtues by standing in various relationships to this common desideratum. Goldman claims that justification is not an entirely separate autonomous form of epistemic value, but rather is derivative from the primary value of true belief. The primacy of "veritistic value" is quite clear in reliabilist theories, but he claims that careful inspection of other traditional approaches, namely, foundationalism and coherentism, reveals a similar dependent status for the value of justification.

Traditional epistemological theories talk of two values, true belief and error avoidance, which seems to point to value dualism rather than value monism. A popular approach to philosophy of science claims that there is an irreducible plurality of cognitive values in science (conservatism, simplicity, generality), not a single value. Goldman challenges both pictures by proposing that a veritistic unity underlies this apparent diversity. Goldman here draws upon adjudication systems in the law and Grice's theory of conversational norms. He concludes that all epistemic values are derived from the value of achieving a high degree of truth possession.

Ernest Sosa's "For the Love of Truth?" begins by raising questions about the following claim: "Rational beings pursue and value truth (the true, along with the good and the beautiful). Intellectual conduct is to be judged, accordingly, by how well it aids our pursuit of that ideal." What does this mean, and is it true? Roughly the first half of Sosa's essay explores a "direct approach" on how to understand the motivation for truth in our intellectual lives and in epistemology: an approach in terms of specific questions and correlated desires for the truth *as such*. Sosa argues that it is at best problematic to claim that rational beings should pursue and value truth, with an interest in the truth *as such*. In the second half, he explores a "more indirect approach," namely, that our truth-connected practice is one that aims not at true beliefs but at truth-conducive practices. On this model, behind every fully justified belief lies a practical syllogism whose main governing principle reflects the practice of aiming for truth. But how are we to conceive of a belief-guiding practice constitutive of our pursuit of truth? We confront a potentially vicious regress. Any hope of stopping the regress, Sosa argues, rests on practices constitutive of our first nature. If your epistemic practices are to be in good epistemological order, this first nature had better be in proper touch with the truth.

But how far back must we go in judging whether *x* is a genuine instance of (occurrent) knowledge? *All the way back?* Enter "Swampman," a being like the rest of us but zapped into existence by lightening. Swampman didn't choose any part of his nature, but isn't he nonetheless an epistemic agent fully capable of having knowledge, if any of us is so capable? Sosa thinks so and concludes that it is not all that important *where* the virtuous habits of epistemic character come from. What *is* important is that the character be stably virtuous.

Abrol Fairweather defends the epistemic significance of motivational character in "Epistemic Motivation." The dominant account of virtue used in virtue ethics is Aristotle's, which makes having an appropriate motivation a component of a state of virtue. The epistemic implication of adopting this view is that knowledge requires having an epistemically appropriate motivation. But two other plausible accounts of virtue — virtues as excellences of faculties and virtues as skills — do not include a motivational requirement. With which general account of virtue should we craft our account of intellectual virtue? If we use the Aristotelian model of virtue in epistemology, then we will require that a believer must have an appropriate motivation — an epistemic motivation — in order to possess intellectual virtue. This would also be a requirement for knowledge according to this kind of virtue epistemologist.

What does it mean to say that a believer has an epistemic motivation? Is this a reasonable condition for knowledge? Fairweather thinks this is a reasonable epi-

stemic requirement and argues that knowledge attributions often depend significantly on the kind of motivational states that direct a person's belief formation. He considers cases where a believer has (1) an inappropriate epistemic motivation (a disregard for truth), (2) no motivation, and (3) an appropriate epistemic motivation (some form of a desire for truth). He argues that the first two believers are sufficiently epistemically defective to warrant a denial of knowledge, but the third believer does possess knowledge. We reach these different evaluations because of the differences in the believers' motivational states, not the kind of evidence they possess. Fairweather concludes that motivational states are epistemically significant and that for this reason the Aristotelian account of the general property of virtue is the preferred account.

Robert Audi's essay, "Epistemic Virtue and Justified Belief," develops the analogy between moral and intellectual virtues and explores its epistemological implications. Like our moral lives, our intellectual lives—our questioning and judging, our reflection and inference, our criticism and responses to others— can be conducted well or poorly. Audi develops the distinction between "epistemic virtue" and "belief from epistemic virtue" by way of examining Aristotle's distinction between virtue and action from virtue and Kant's distinction between action from duty and action according to duty. Armed with this distinction and its history, Audi defends a virtue-based analysis of knowledge, epistemic responsibility, and justification over a more particularist, belief-based analysis.

Heather Battaly examines the possibility of fruitful debate within VE in "Thin Concepts to the Rescue: Thinning the Concepts of Epistemic Justification and Intellectual Virtue." She begins by examining the arguments of William Alston and Stewart Cohen, which purport to show that many contemporary debates over justification, for example, the internalism/externalism debate, are fruitless because there is no single concept about which the parties to the debate disagree. She notes that similar debates are emerging in VE. Sosa and Goldman use the concept of intellectual virtue to ground some form of reliabilism, while Linda Zagzebski and James Montmarquet use the concept of intellectual virtue to ground either some form of internalism or a hybrid theory. If Alston and Cohen are correct, then it appears likely that VE is set for the same kind of fruitless debates that have characterized much recent epistemology.

Battaly introduces the notion of a "thin concept" to rescue the fruitfulness of epistemic debate. A concept is thin, according to Battaly, if only some conditions of its application are fixed, or garner agreement among competent speakers, leaving a range of properties that are not universally acknowledged as either necessary or sufficient conditions for falling under the concept. Thinness is a matter of degree for Battaly; the more conditions of application that are fixed, the thicker the concept. Rather than seeing internalists and externalists as working with different concepts of justification, she argues that they are better seen as employing a common but "thin" concept of justification that is thickened in different ways by different theorists. She argues that identifying a thin concept of justification blocks Alston's and Cohen's arguments for the fruitlessness of epistemic debates and preserves the significance of the emerging debates within VE.

In "Virtues and Rules in Epistemology," John Greco argues that virtue theories in epistemology hold an advantage over deontological theories in epistemology because the former need not understand epistemic justification in terms of epistemic rules or norms. Greco begins defending his claim by looking at action-guiding *moral* rules: it is an advantage of virtue theories, it is argued, that they do not require that moral action be understood in terms of such rules. This essay argues in a similar way with respect to theories of epistemic evaluation. The argument against deontological theories proceeds in two parts since there are two major kinds of deontological theory that must be refuted in different ways.

Weak deontological theories hold that one's belief is justified so long as it does not violate any relevant rule. The main objection against such theories is that they fail to take into account the causal etiology of belief. For example, such theories fail to distinguish between having good reasons and believing for good reasons. But whether a belief qualifies as knowledge is partly a matter of etiology, and so weak deontological theories are too weak. Strong deontological theories hold that one's belief is justified only if it results from following the rules, as opposed to merely being describable by them. The main argument against strong deontological theories is that it is an empirical question, concerning a contingent matter of fact, whether human cognition is governed by rules. In contrast, virtue theories make causal etiology matter, requiring that in cases of knowledge belief is the result of virtuous cognitive character. However, virtue theories need not require that knowledge be governed by rules. On the contrary, they can make this an empirical question about the mechanisms of human cognition rather than a philosophical question about the conditions for knowledge.

Linda Zagzebski pursues the issue of how epistemic evaluation depends upon human agency in "Must Knowers Be Agents?" In particular, she raises the following questions:

1. What are the conditions for being an effective agent? What determines that an agent is effectively exercising her agency on a particular occasion? Must she be reliable? Is her efficacy determined by what she is able to do in counterfactual circumstances?
2. Is there any important difference between an effect arising from the act of an agent, whether voluntary or non-voluntary, and events brought about by a non-agent? In particular, does it make any significant difference to epistemology?
3. Is knowledge best understood on the model of event causation or on the model of agent causation?

As an aid to answering these questions, Zagzebski proposes epistemic analogues to so-called "Frankfurt cases" against the Principle of Alternate Possibilities: A person is not morally responsible for an act unless she could have done otherwise. Zagzebski argues that the moral of Frankfurt cases is that manipulable counterfactual conditions are neither necessary nor sufficient for either moral or epistemic responsibility, nor are they necessary for knowledge. But it would be a mistake to conclude that they are irrelevant. Typical counterfactual conditions are signs of what

really *is* essential to responsibility and knowledge — the presence of agency. Zagzebski concludes with a discussion of a problematic case for her agency view: the simplest perceptual beliefs. She argues that agency is preserved in the possession of even the simplest perceptual beliefs, and we can have knowledge in such circumstances, as long as they are endorsed by the reflective mind.

In "Epistemic Luck in Light of the Virtues," Guy Axtell identifies the development of "mixed" externalist epistemologies as a shared project among contemporary virtue epistemologists. This is an account that is generally externalist in character, yet which blends its objective "success" conditions on warrant or justified belief with subjective "responsibility" or motivation conditions on epistemic agents. Yet despite this widely shared goal, virtue epistemologists disagree widely on a number of important issues pertaining to defining the intellectual virtues, responsibility for character, the strength of the analogy between ethical and epistemic evaluation, and the prospects for a unified account of the virtues. Axtell uses their responses to the problem of "epistemic luck" as a sounding board and locates the source of these disagreements in divergent, value-charged "interests in explanation," which epistemologists bring with them to discussions of knowledge and justification. In so doing, he delineates both the commonalities and key differences between those authors he describes as *virtue reliabilists* and those he describes as *virtue responsibilists*.

In his analysis of epistemic luck, Axtell shows how "unmixed" internalist and externalist epistemologies must each acknowledge a different form of epistemic luck as a consequence of their theoretical approach, leaving each open to devastating criticisms by their adversaries. From this Axtell concludes that neither approach, freestanding, is adequate to respond to the challenge of skepticism. Finally, Axtell considers and responds to objections to his analysis, focusing especially on the serious charge that mixed accounts are "compromises" and, by their very nature, philosophically "unstable." Axtell seeks to undermine this objection by showing in detail how mixed accounts — particularly those that utilize the resources of virtue theory — are better able to respond to the challenge of skepticism.

Christopher Hookway's essay, "Epistemic *Akrasia* and Epistemic Virtue," explores the concept of *akrasia* (incontinence, or moral weakness) in epistemic contexts. In the practical realm, studying forms of irrationality such as *akrasia* can provide important clues to the psychological structure of rational behavior and the kinds of evaluations we have to make when we try to act well. Hookway's essay explores the possibility of learning similar lessons from studying epistemic *akrasia*, a phenomenon that would involve our believing what we know we are epistemically wrong to believe. Hookway argues that epistemic rationality depends upon the possession of states that govern inquiry in the way in which virtues are held to govern practical reasoning and that "continence" is a fundamental executive virtue.

It is unsurprising that inquiry and deliberation can exhibit practical *akrasia* because they are goal-directed activities: I can knowingly inquire or deliberate in ways that conflict with my standards of good inquiry. Hookway distinguishes between standard *akrasia* of the sort described above and full-blooded *akrasia* that involves believing a proposition when I know there to be a very strong reason to believe its negation. Many philosophers have argued that full-blooded epistemic *akrasia* is in-

coherent because having sufficient reason to believe *p* and believing *p* are too intimately related to come apart in the way full-blooded *akrasia* requires. Hookway argues that once we pay close attention to the fine structure of the ways in which propositions and arguments become salient and influence our attention, we can see that full-blooded epistemic *akrasia* is possible. These reflections lead Hookway to an account of epistemic virtue that gives pride of place to the virtue of "epistemic continence."

Keith Lehrer identifies the role that intellectual virtue plays in discursive knowledge in his contribution, "The Virtue of Knowledge." Discursive knowledge is the kind of knowledge that a subject can use as a premise in reasoning to confirm some conclusions and reject others. As a coherentist, Lehrer interprets justification in terms of a belief's coherence with a background system that is undefeated or irrefutable in terms of errors in the system. Lehrer argues that attaining justification in this sense requires that a person exercise intellectual virtue in accepting what she does. When intellectual virtue in what a person accepts explains why the person succeeds in obtaining the objective of truth, the justification based upon coherence with the background system will be undefeated and convert into knowledge. According to Lehrer, the success of virtue yields the virtue of knowledge. His essay also seeks to clarify some tensions between subjective and objective, internalist and externalist, as well as motivational and reliabilist approaches to the subject of virtue by considering the role of intellectual virtue in knowledge.

Richard Foley focuses on the concept of epistemic responsibility in "The Foundational Role of Epistemology in a General Theory of Rationality," but he seeks to illuminate this important epistemic concept from the perspective of a general theory of rationality—a theory that addresses the rationality of actions, policies, and plans as well as beliefs. A common complaint against contemporary epistemology is that its issues are too rarified and, hence, of little relevance for the everyday assessments we make of each other's beliefs. The notion of epistemic rationality focuses on a specific goal, that of now having accurate comprehensive beliefs, whereas our everyday assessments of beliefs are sensitive to the fact that we have an enormous variety of goals and needs, intellectual as well as non-intellectual. The latter, Foley argues, have an ethical or quasi-ethical dimension: We want to know whether someone has been responsible, or at least non-negligent, in forming opinions. Nevertheless, epistemology, properly conceived, is relevant to our commonplace intellectual concerns. The epistemologist's notion of epistemic rationality, while an idealized notion, serves as an anchor for the general theory of rationality that we use in our everyday assessments. By properly locating epistemic rationality within the general theory of rationality, Foley argues that it is made more relevant to our assessment of the rationality of beliefs.

Hilary Kornblith examines the role of empirical research in theories of epistemic obligation in "Epistemic Obligation and the Possibility of Internalism." Some have argued that the very idea of epistemic obligation presupposes doxastic voluntarism. Richard Feldman has argued that it does not. But Feldman wishes to use his defense of the legitimacy of epistemic deontology as a springboard for a particular account of our epistemic obligations, a variety of internalism. Kornblith argues that Feld-

man's defense of the legitimacy of epistemic obligations does not leave room for his defense of internalism. In the end, Feldman's views about epistemic obligation suggest an altogether different defense of internalism than the one that he himself wishes to endorse. Kornblith argues that the most defensible form of internalism is committed to an interesting and controversial empirical research program.

3. The Future of Virtue Epistemology

In comparison with virtue ethics, VE is still in its infancy. Epistemologists are only beginning to take seriously the idea of an intellectual character, as well as such attendant notions as intellectual motive, end, agency, and freedom. The connection of VE with virtue ethics raises a wide range of new questions, some of which have not yet even been mentioned in print. Some unexplored or barely explored questions include the following: Which of the many notions of virtue is best suited for epistemic evaluation? Are the moral and intellectual virtues unified? Should we investigate epistemic psychology as the analogue of moral psychology? What is the proper place of emotion and other affective states in the acquisition of knowledge? What relation must affective states bear to doxastic states in order to confer epistemic praise? Are certain affective states intrinsically praiseworthy or only insofar as they bring about true beliefs?

Other new questions parallel standing discussions in ethics: How do we balance epistemic principles and virtue in our theory of knowledge? Does the distinction traditionally used in ethics between subjective and objective duty have a parallel in epistemology? What is the connection between practical and theoretical rationality? What is the connection between meta-ethics and meta-epistemology? Are there ultimate human ends, and if so, how are they connected to our epistemic ends? Are there important epistemic ends other than knowledge and rational belief? What is the place of one's epistemic community in the acquisition of evaluatively positive epistemic states? What about epistemic vice? The focus of attention in VE, as in virtue ethics in general, has typically been on evaluatively positive traits in spite of the fact that the negative traits are surely more common. But we do not necessarily understand vice by understanding virtue. In fact, there are many distinct evaluative levels in addition to virtue and vice, as Hookway's essay on epistemic *akrasia* demonstrates. If there is a difference between epistemic vice and epistemic incontinence, is there also a difference between epistemic virtue and epistemic continence? These questions and many others deserve attention.

In addition to these new questions, some old ones can be given a different spin when approached from the standpoint of VE. One is the question of whether VE dissolves the internalism/externalism standoff. Virtue is a complex and forgiving norm, and this allows it to fill a number of theoretical needs in epistemology. It happens that the dispute between internalists and externalists can be framed nicely within a virtue-based framework. Since some virtue epistemologists maintain that both the causal history and efficacy of a person and her motivational states are important in conferring virtue, both internalist and externalist requirements must be satisfied in the pos-

session of virtue. Rather than ending up with radically opposed epistemic theories, this form of VE proposes a unified framework that admits the value of both criteria. Another old question is the foundationalist/coherentist dispute. VE does not come down on one side of this controversy, but gives a method for deciding it: the behavior of intellectually virtuous persons. As we've already said, some of these old issues may deserve less attention than they've received for the past several decades, but the oldest question of all is the central one in epistemology and that also gets a different answer when approached from the side of virtue: the nature of knowledge.

VE is an exciting field of inquiry in part because of the way it raises questions that overlap with the concerns of other fields of philosophy. Besides interfacing with ethics, work on epistemic psychology arising from VE is likely to merge with the new interest among Anglo-American philosophers in philosophy of emotion. In addition, VE has the potential to go much farther than traditional approaches toward incorporating the social dimension of knowing. That is because the acquisition and exercise of virtue requires a rich social bedding; knowledge so conceived reaches beyond the individual knower into his social environment. VE therefore conforms nicely with the emerging field of social epistemology.[15]

The essays in this volume raise a multitude of questions that deserve more detailed exploration. We encourage epistemologists—as well as philosophers working in ethics, philosophy of mind, action theory, and social philosophy—to investigate the many issues emerging from a virtue approach to epistemology.

Notes

1. E. Gettier, "Is Justified True Belief Knowledge?," *Analysis* 23 (1963): 121–23.

2. See K. Lehrer and T. Paxson, "Knowledge: Undefeated Justified True Belief," *Journal of Philosophy* 66 (1969): 225–37; reprinted in G. S. Pappas and M. Swain (eds.), *Essays on Knowledge and Justification* (Ithaca, N.Y.: Cornell University Press, 1978). Also Marshall Swain, "Epistemic Defeasibility," *American Philosophical Quarterly* 11 (1974): 15–25; reprinted in Pappas and Swain, *Knowledge and Justification*.

3. A well-known recent example of a bifurcated theory is developed in Alvin Plantinga's *Warrant The Current Debate* (New York: Oxford University Press, 1993).

4. See William Allston's "Epistemic Desiderata," *Philosophy and Phenomenological Research* 53, 3 (Sept.): 527–51.

5. For examples of such arguments, see Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979) and Michael Williams, *Unnatural Doubts* (Princeton: Princeton University Press, 1996).

6. The historical connection between skepticism and justification is discussed by Linda Zagzebski in "Recovering Understanding," *Knowledge, Truth, and Obligation*, Matthias Steup (ed.) (New York: Oxford University Press, forthcoming).

7. "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge," in *Studies in Epistemology: Midwest Studies in Philosophy*, vol. 5. (Notre Dame, Ind.: University of Notre Dame Press, 1980).

8. Ernest Sosa defends faculty reliabilism in a number of essays collected in *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991), the earliest of which is "The Raft and the Pyramid."

9. Alvin Goldman defends process reliabilism in "Epistemics: The Regulative Theory of Cognition," *Journal of Philosophy* 75 (Oct. 1978): 509–23.

10. Roderick Chisholm, *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice Hall, 1966; 2nd edition, 1977; 3rd edition 1989).

11. See Code's *Epistemic Responsibility* (: University Press of New England for Brown University, 1987), and Montmarquet's *Epistemic Virtue and Doxastic Responsibility* (Lanham, Md.: Rowman and Littlefield, 1993).

12. "Agent Reliabilism," *Philosophical Perspectives*, 13 (*Epistemology*), (Atascadero, CA: Ridgeview Press, Fall 1999). Greco uses the term "agent reliabilism" for a larger class of theories than his own, including Sosa's, Plantinga's, and my early theory.

13. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (New York: Cambridge University Press, 1996).

14. For more on how VE deals with skepticism, see Greco's book, *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry* (New York: Cambridge University Press, 2000).

15. See Alvin Goldman's *Knowledge in a Social World* (New York: Oxford University Press, 2000) and J. Angelo Corlett's *Analyzing Social Knowledge* (Lanham, Md.: Rowman and Littlefield, 1996).

2

REASON, VIRTUE, AND KNOWLEDGE

Simon Blackburn

1. Setting the Scene

My aim in this essay is first to clarify what any position worth calling "virtue episte-mology" ought to hold. I then want to explore some of the relations between such an approach to epistemology and two other doctrines. One is a minimalist or deflation-ist conception of truth. The other is a generally expressivist approach to values and virtues, and hence to rationality.

It is, I believe, a very attractive idea to take what can be said about moral virtue and see how it looks when applied to intellectual or cognitive virtues. If truth or per-haps knowledge or wisdom is the goal of intellectual endeavour, then it might be re-garded as playing the parallel role to *eudaimonia* as the goal of living. And then we should expect any account of the traits necessary to achieve the one as quite strictly parallel to the account of the traits, the virtues, necessary to achieve the other.

Furthermore, there are some fairly immediate points of contact. Fair-minded-ness, courage, judgment, and experience can be involved in the cognitive domain just as they are in the practical domain. We might reflect, as well, that faults in the cognitive domain, such as that of being too timid or too stubborn or insensitive or prone to fantasy, would directly reflect, or indeed be part of, wider moral faults. And on some accounts of ethics, all moral faults are at bottom not only analogous to cognitive faults, but are actually identical with them. If to know the good is to love it, then moral defect becomes a species of cognitive defect. And it could in return be suggested that many cognitive defects are at bottom moral and that only cognitive defects that are beyond our control, such as those caused by unavoidable external or internal obstacles to inquiry, fail to qualify as moral defects.

However, if virtue epistemology is modeled upon virtue ethics, then I think we need more than these relatively straightforward points of contact. In particular, I suppose that, like virtue ethics, if it really is a distinct approach to ethics, virtue epis-temology will need to defend a certain kind of priority. Consider the following equations:

(1) An action produces (or tends to produce, or is such as to produce) the greatest balance of benefit over harm of any alternative if and only if it is the action that would be performed by a virtuous agent.

(2) An action is the right action to perform in the circumstances if and only if a virtuous agent would perform it in the circumstances.

Some people attached to these equations might advance them as undercutting anything distinctive about virtue ethics. The equivalences, in other words, give us a fix on what is true of a virtuous agent, and that is all. So the consequentialist critic would be supposing that we have, antecedently, a conception of the balance of benefits over harms, and in the light of that we can use the first equivalence to define what the virtuous agent does. We might suppose that we have an independent grip on what it is to be happy, just as we have with regard to pain and misery. Then the promotion of one and diminution of the other is indeed a self-standing aim, understood independently of virtue and available to act as at least one test for when a quality is indeed a virtue. The deontological critic would similarly say that we have, antecedently, a conception of the right action to perform in given circumstances, and read from the second equivalence that this is what the virtuous agent does. Thus we might hold that some such test as Kant's gives us an entrée into the notion of the right, after which we can indeed select as virtues traits that gain expression in right behavior. The *Oxford Dictionary* supposes this, defining virtue in the moral context as "Conformity of life and conduct with the principles of morality; voluntary observance of the recognized moral laws or standards of right conduct; abstention on moral grounds from any form of wrong-doing or vice." Here the concepts of wrong-doing and vice come first, and virtue is understood in terms of them.

The virtue ethicist can respond, of course, by denying the equations outright, either in these simple forms or in any more complex forms. The more interesting reaction is to accept them, but to read the equations the other way round, or "right to left." She will say that we have a conception of what virtue would have us do, and in the light of that we fashion our concept of the balance of benefit over harm, or a concept of what it is right to do.

So, for instance, when Hume says that personal merit, or virtue, consists in the presence of qualities "useful or agreeable to ourselves or others," this type of virtue ethicist need not disagree. But she has to insist that this does not amount to a definition or explanation of what a virtue is in terms that can be independently understood. She insists, instead, that our concept of what is useful or agreeable is partly or wholly derivative from our conception of what living virtuously requires. If Hume intended his formula as a consequentialist *account* of how some trait gets to be on this list of the virtues, then this virtue theorist disagrees, insisting instead that whatever truth there may be in the account presupposes an independent conception of virtue. A key element in the virtue ethicist's response will be that "usefulness" or "agreeableness" or, more generally, happiness itself is to be understood primarily in terms of living virtuously. If this seems too pious, then perhaps the claim will be that while false or hollow happiness may coincide with failures of virtue, this is what real

or true happiness consists in. This, supposedly, is why Aristotle's invocation of *eudaimonia* is not a consequentialist departure from single- minded virtue ethics.

In the recent literature, some philosophers seem to want to call themselves virtue epistemologists without accepting the priority I have identified. They may want to allow the priority of other notions, notably that of truth, and simply confine themselves to emphasizing the value of an alignment of belief and truth, or of the traits that contribute to that alignment. Or, they may want to insist that "everything comes at once," so that there is a circle of terms none of which can be understood antecedently to the others. For the purpose of this essay I do not want to legislate. We can simply distinguish strong virtue theories, which hold the "right to left" priority, from weak virtue theories that have no such commitment. The problem for weak theorists will be that of finding a distinctive voice, enabling them to distinguish themselves from simple reliabilists.[1]

The strong virtue theorist's priority is, of course, surprising to some. In the moral case, it might seem to face the disadvantage that it leaves no account of why a quality *does* get on the list of virtues. The standard Aristotelian move against this objection is to cite the parallel with flourishing in plants and animals. We know what it is for a primrose or a tiger to flourish, and the same is supposedly true of ourselves. The virtues then become those traits that make up or contribute to human *eudaimonia*. This flirts with the same danger as presented by utilitarianism, which is that all the work is done by the idea of promoting flourishing, with the virtues just tagging along. So in order not to collapse into a kind of utilitarianism the virtue theorist needs not only that human flourishing is strictly analogous to animal or plant flourishing, but also that it distinctively includes certain ways of acting (justly, charitably, and so forth). It is not at all obvious that the combination is stable.[2] Trees often flourish by making life impossible for other trees, and the same seems to be true of human beings. But this is not our present concern.

How does the parallel dialectic emerge in epistemology? What would the equivalences parallel to those above look like? I am going to suggest three. One is concerned with probability or justification, a second with knowledge, and the third with truth.

(3) A proposition is probable (justified) in a circumstance *C* if and only if an epistemically virtuous agent in *C* would have confidence in it.

(4) A true proposition is known to be true by an agent *S* in circumstance *C* if and only if *S* in *C* exhibits epistemic virtues in accepting it.

(5) A proposition is true if and only if an epistemically virtuous agent would accept it, if he exercised the virtues appropriately.

Read right to left, these are in ascending order of ambitiousness. It is not so very radical to associate probability or justification with a virtue, such as rationality in distributing confidence. It is probably more radical to think of capturing knowledge in a similar way, and most radical to aim at the concept of truth itself.

Clearly, as they stand each of these is very rough and could be refined much further. For example, (3) could be given a more quantitative formulation, matching de-

grees of probability to degrees of confidence. (4) would need refinement to protect against the fairly obvious counterexamples deriving from misleading circumstances in which virtue leads the epistemic agent astray and so on. (5) would need similar refinement, perhaps leading in the direction of Peirce's conception of what virtue (in his hands, scientific method) would lead to if pursued in some presumed long run. (5) would also need some work to make it relate satisfactorily to (4). The difficulty is that if truth is described in terms of what a virtuous agent would accept, knowledge cannot be similarly defined on pain of eliminating the distinction between the two. This can be seen because if we try substituting the equivalence in (5) for the occurrence of 'true' in (4), we seem close to collapsing knowledge and truth. There is some space, however, between what a virtuous agent *does* accept, which is what is mentioned in (4), and what he or she *would* accept, which determines (5). The combination would deliver the idea that truth is what you would get to by investigating virtuously, whereas knowledge is what you have got when you *have* investigated virtuously. Whether this is exactly the right gap between truth and knowledge is clearly disputable. Perhaps its only merit is that it does at least reflect the idea that there is normally no gap between aiming at knowledge and aiming at truth. A final qualification concerns the "circumstances" mentioned in (3) and (4): What a virtuous agent would accept will often not depend upon external or objective circumstances so much as upon her internal theories and beliefs, or the circumstances insofar as she is capable of appreciating them. Circumstances are, as it were, intentional.

However doubtful or attractive the equivalences are, there is still the lurking question of priority, and it is this upon which I want to focus. Just as with ethics, there will be theorists who suppose that even if the equivalences can be spruced up, they merely tell us what epistemic virtue requires, given antecedent conceptions of knowledge, probability, or truth. Virtue would be identified in terms of aligning our beliefs with the truth, which is why (5) is more or less plausible. Justification means adjusting our confidences to probabilities, explaining (3). And knowledge arises when we accept propositions in circumstances that require their acceptance, which explains what is right about (4). Read like this, the equivalences are too weak to suggest any distinctive approach to epistemology. A virtue epistemology this weak is only a fig leaf for reliabilism.

2. Justified True Belief

Clearly the equivalence (4) is close to the familiar "justified true belief" (JTB) account of knowledge, and with some versions of the refining I suggested in the last section, would quickly turn into it. And then the question of priority is certainly on the table, with classical JTB theorists claiming that knowledge is what you get when your true beliefs are justified, and rivals claiming that justification is only identifiable as that which turns true belief into knowledge. Here, the JTB theorists are the virtue theorists, since they take the notion of virtue or justification as prior to that of knowledge, which is to be described or defined in terms of it. The rival priority sees

justification as itself only identifiable in terms of a prior conception of "whatever it takes" to turn true belief into knowledge, here taken as the primitive.

But (4) need not be refined in just that way, and there are issues at stake in so treating it. Everything will depend on how the notion of a virtue maps onto the notion of a justification. It is indeed epistemically virtuous in some cases to be able to *produce* justification for a particular proposition p, by citing supporting propositions $q, r \ldots$ And it is virtuous in more cases to be able to *recognize* such justifications when they are provided. But it ought to be highly contentious to claim what I would regard as false, namely that epistemic virtue is *exhausted* by such abilities. One thread in the meaning of "virtue" is just that of a power or efficacious quality, and it is quite open to us to privilege other powers than sensitivity to relations of propositional confirmation. One virtue we like in guides and informants is the ability to *get things right*, or sheer reliability. And reliability cannot be reduced to sensitivity to confirmation relations, for two reasons. First, such sensitivity is not sufficient for reliability: At the very least, it presupposes that the evidential propositions are reliably believed. And second, it is not necessary, because reliability given by perceptual mechanisms and memory is not a matter of sensitivity to evidence and inference.

I suspect that philosophers have been slow to recognize the need for both elements because of combative labels like "externalism" and "internalism," with the implication that there is a single choice to be made. The externalist then insists on the way knowledge or justification depends upon whether, perhaps fortuitously, we have the right relations to the realities we are describing. The internalist stresses the need for right reason in handling the inferential relations among the descriptions. The obvious, peaceable remark is that the well-tuned agent needs both. It is absurd to see a happy relationship to the reality as any kind of *rival* to sensitivity to propositional confirmation. It is a complementary part of what makes up epistemic virtue. At first sight reliability is of more concern in cases like direct perceptual awareness; sensitivity to confirmation relations is more immediately visible in the scientist or the detective or the judge.

In his exploration of the concepts of experience and justification, McDowell cautions us against a tempting dualism at this point. "Experience," we might think, is one thing; propositional justification, or justification "within the space of reasons," is another. If we think like this, he warns us, the contribution of the world to our thinking will be a "brute impact from the exterior," and such brute impacts, while they may *exculpate* our arriving at some beliefs, cannot *justify* those beliefs: "in effect, the idea of the Given offers exculpations where we wanted justifications."[3] The idea, I take it, is that we are not to blame if, as the recipients of some brute impact from the exterior, we end up thinking whatever we do, under the causal influence so provided. But neither have we entered the realm of justification: McDowell's comparison is with someone swept to a place by a tornado, who is then neither justified nor to blame if significance attaches to his being there. But, if it is essential to our self-conception that we are justified even when we form simple perceptual beliefs, an account of what is going on that cannot deliver that is thereby refuted.

Whatever may be wrong with talk of a "Given," it is hard to believe that this diagnosis reveals it. Responding to a causal impact by coming to believe something about its origin, it seems to me, is not just something that could stand as an "exculpation" when things are going wrong. Done as a habit, it shows the agent to have a virtue, a power or ability to get things right (being swept along by tornadoes shows no such virtue). Consider the familiar case of proprioception. Here, under the "brute impact" of postures of our bodies, themselves causing signals in muscles and nerves, we acquire the ability to judge, unhesitatingly and rightly, where our hands or feet are. That ability is a virtue — someone without it suffers from a lack or deficiency, and one that could imperil their health or survival. We may, if we wish, say that the belief that my hand is at the back of my head "lacks justification," meaning that I can offer nothing to say why I believe it, except that it is true. But that does not stop us from saying that in forming it as I did, I exercised a distinctive virtue, indeed arguably the cardinal epistemic virtue, namely that of getting it right.

Although he makes much of it, I do not think the issue over justification is central to McDowell's overall project of showing that conceptual capacities are drawn on "in" experience. So far as I can see, the point that the right kind of receptivity to "brute impacts" shows a virtue is quite compatible with saying that even the most primitive level of experience or consciousness is already partly the work of conceptual capacities. This larger doctrine has its own attractions, although it gets into notorious difficulty with animal experience. It is certainly plausible to think, in the case of proprioception, that the first thing, as it were, that enters consciousness is an awareness *that*, rather than some more basic experience that just sits there waiting to be taken one way or another. But the attractions of the doctrine will have to be put a different way. For there need be nothing wrong with the idea that a correct response to a "brute impact" or a brute given exhibits virtue. This is so whether or not the impact or the given is now thought of as itself determining an element of consciousness, or regarded simply as a causal element reliably giving rise to the (conceptual) denizens of consciousness, such as judgments.

When an agent exhibits virtue we can perfectly well say that *she* is justified. It is not that her *belief* is justified by a different belief. It is just that this sensitivity to the way of things is exactly what justification amounts to; for example, when you respond correctly to your hand being behind your head. The same is true when you recognize your friend or a voice, or the spatial configuration of the landscape around you, or in any other simple perceptual case.

I should mention as an aside that once this point is taken, it becomes quite unclear what is achieved by the chorus of complaints about "the myth of the given." There are, we know, causal processes that end up with my thought that there is a cardinal in the garden, because I can see it; that the cat is at the door, because I heard it; or that dinner is curry tonight, because I can smell it. Modern critics of the given want to insist that the causal process does not work by intruding into consciousness an unconceptualized sensation, a kind of qualia, that is then interpreted or understood, ending up with the belief or thought.[4] They may be right. But why is it so *important* whether it works that way, or in the other way, whereby the first element of consciousness is itself conceptual? Do we know enough about consciousness to re-

gard this question as so utterly seminal? In terms of the familiar metaphor, the question is whether the field of consciousness is entirely "the space of reasons" or whether it may be the space of reasons and something else, such as traditional sensations or qualia. I can see that the question has its own interest, and part of that interest is that it is surprisingly difficult to pronounce upon. But it is not really at all plain what difference it makes to epistemology, nor even that it is pivotal in destroying empiricism or foundationalism.

In fact, so far as foundationalism is concerned, "qualia" do not help, for even if we countenance them, there is no reason to suppose that they compel belief and judgment. The art of recognizing a smell, for example, might be best thought of in coherentist ways even if smells themselves are elements of consciousness that are indeed just given. I mention this only as an aside.

Reliability, then, is an epistemic virtue, just as knowing his way around is the cardinal virtue in a guide. But McDowell's comparison with being swept away by a tornado nicely illustrates the discomfort some people feel here. It introduces an element of externality: How reliable we are may not be entirely under our control, but partly a gift of, for example, a friendly and familiar environment. And being out of control strikes us as unworthy and bad. This is parallel to the familiar idea in ethics that our moral virtue must be entirely within ourselves—a matter of how we will—and not a matter of external relations, or brute happenstance or luck, whereby our actions turn out well. I am not impressed by this thought, which seems partly to depend on an unsustainable metaphysics of free will. But more important in this context, it depends on forgetting that our external relations are themselves matters that we can register and control, and matters that we may be blameworthy for mishandling or neglecting, however well-meaning we may be. In epistemology, reliability is *partly* a gift of nature (the blind are not reliable over colors) and may *partly* be a gift from a friendly and familiar environment, but it is also something we can monitor and improve and manage, and therein lies our responsibility. The comparison with being swept away by a tornado is quite wrong. The reliable perceiver exercises his virtue partly by knowing which causal impacts to put himself in the way of. You can control where you go and whether you open your eyes when you go there, and you can exercise judgment partly by recognizing that your situation is too impoverished for judgment to be warranted.

A difficulty now confronts our epistemic virtue theorist. If, as I believe, reliability sits firmly in the center of cognitive virtues, then the priority she needs to defend seems to be reversed. For there *is* an account of how a trait gets to be on the list of epistemic virtues. It will be there because it promotes an alignment of belief and truth. This is parallel to the criticisms of virtue ethics displayed above; it is like saying that a trait counts as a virtue because it promotes utility or wards off loss, where utility and loss are independently understood. Yet it seems difficult to imagine epistemology without this account.

3. Truth and Virtue

Reliability is naturally defined in terms of truth. The virtue is that of cleaving to the true and avoiding confidence in the false. So it is natural to worry that only a robust or thick conception of truth will sustain our sense of the virtue.

This worry can be illustrated. Indeed, perhaps it is all too obviously illustrated by the entire climate of "post-modernism," which, having convinced itself that talk of truth could only be some kind of fraud or mask for power or whatever, rapidly lost any respect for any particular way of conducting historical or intellectual or perhaps any other kind of inquiry. Intellectual processes become evaluated in other terms. The virtues of the inquirer are no longer those of reliability or accuracy or ability to marshal evidence and compel belief, but other things entirely. In light-hearted versions, the virtue becomes that of cutting an agreeable figure in the carnival. In more sombre versions, it becomes that of pursuing one or another political or religious agenda.

We can illustrate the problem by considering Peirce's conception of truth as that which would be agreed upon in the limit of scientific investigation. For this to work, it seems, we need a satisfactory conception of the value that attaches to scientific investigation, other than that it is the midwife to the truth. Without that, we lose any conception of the respect such investigation deserves, or of the difference between conducting it properly or improperly. If we cannot conceive of a process as organized in a virtuous direction, we cannot respect any point on which it might converge. In other words if the end product, truth, does not confer value upon the processes that reveal it, then the processes themselves must carry a merit that they confer upon the end. Yet it will be difficult to say just what is good about some methodology or another except that it makes for truth. Compare, for instance, the virtues of the historian with that of the novelist. How could we sustain any conception of how it is *right* for each to conduct themselves, and why it is right that in some respects they conduct themselves differently, without recognizing that the historian is answerable to real events in a way in which the novelist is not? Each may be doing something enjoyable or political or difficult or imaginative or gripping, but the historian is doing something else as well, and without understanding that we cannot understand the virtues of the process. And it seems impossible to imagine an understanding of that "something else" that does not explicitly or implicitly identify it as the concern to find truths about the past. Without knowing that, we would not know what game is being played.

If this is indeed the situation then, as just threatened, virtue epistemology will not be able to defend its priority. And it may look as though we need a robust or thick conception of truth if we are to justify the alternative priority, where truth is valuable and intellectual traits are classified as virtues insofar as they lead to it.

I believe, however, that this second point at least will not stand. For I take it that a minimalist or deflationist theory of truth can, perhaps surprisingly, deliver the requisite sense of the value of truth. The deflationist I shall consider believes that our understanding of truth is simply manifested in our disposition to accept instances of the schema "proposition p is true if and only if p."[5] He adds an account of the value

of the term, given that modest role, which is typically in terms of framing general-izations and disjunctions. Thus, "something Fred told me was true" comes out as summarizing the open-ended disjunction of conjuncts: "either Fred told me p and p, or Fred told me q and q, or. . . ." I shall not describe further the development of deflationism, since it has been admirably done in the work cited.

How is the *value* of truth expressed given deflationism? Consider the schema:

(D) It is good that, if p then I believe that p; and it is good that if I believe that p, then p.

If we are disposed to assent to instances of this schema, then we hold, for example,

(D$_1$) It is good that, if cheese is in the refrigerator then I believe that cheese is in the refrigerator; and it is good that if I believe that cheese is in the re-frigerator, then cheese is in the refrigerator.

The good is described in terms of two conditionals. But neither of them mentions truth. Yet, it may plausibly be claimed, the disposition to assent unrestrictedly to these conditionals is just the disposition to value truth.

Of course, that is consistent with holding that it is more important to satisfy some instances of the conditionals than others. When the p in question is highly signi-ficant, it is more important than when it is not. More merit attaches to some discov-eries than others. Some truths are more important to the historian or the scientist than others. But this does not stop truth from being a value. It just means that it is not the only value.

If we go on to ask why we should be disposed to hold these conditionals, a vari-ety of approaches may be tried. Pragmatism and adaptive explanations take us some of the way. Some philosophers, notably Stephen Stich, have doubted these explana-tions, pointing to individual cases in which false belief stands you in good stead: Mis-taking the time of your flight, you avoid the airplane crash.[6] This kind of argument is not very compelling. We might suppose that such cases are necessarily exceptional, parasitic upon a general alignment of belief and fact. And in any event one prophy-lactic for the calamity that true belief in the airplane's time would bring upon you is to have even more true belief, including belief about whether the plane is in danger. Such true beliefs on the part of the mechanics and the pilot would have worked even better.

More interesting examples come from the adaptive advantage of systematic dis-tortion. The animal that constantly overestimates the chances that a rustle is a pred-ator may survive better than one that estimates the chances rightly. In such situa-tions, "quick and dirty" habits of belief formation may work better than precise and discriminating ones. While I think there is more to be understood here, I don't see there as being scope for a general critique of the value of satisfying instances of (D). In any event, the fact is that we are curious, we dislike ignorance, and we dislike liv-ing in fools' paradises. We value satisfying (D) as often and as fully as life permits.

How do cognitive virtues look on this picture? A virtue will be a trait or dispo-

sition that, when exercised, typically increases the chances of these conditionals being true. Thus, compare three policies for forming confidence that there is cheese in the refrigerator. One is to consult memory and remember buying cheese, say, a week ago. Another is to discount memory but to go and look. A third is the joint policy of going and looking in light of what you remember. Given the nature of human perception and memory, it is plausible that the third maximizes your chance of satisfying the conditional. Unaided memory may mislead you, and in any case someone may have eaten the cheese in the interim. Unaided perception may glance over the place the cheese is hiding. Perception exercised in reasonable confidence that there is cheese there to be found, that is, the third strategy, may lead you to do more than barely glance and thereby minimizes the chance of D_I being false.

Consider now a different cognitive trait. Suppose someone whose background experience gives her very quick intuition in some area: Imagine a doctor quick at diagnosis, or a judge of character quick at reading it from visible gesture or facial configurations invisible to others. Is such intuitive speed a cognitive virtue? Surely the question hinges entirely on whether the diagnoses are borne out. The doctor can be as intuitive as she likes, but if she constantly misdiagnoses patients her speed and sureness become vices. The judge of character may have an enviable speed and certainty, but not if she constantly misreads the signals. Again, the question of whether a trait is classed as a virtue or a vice hinges centrally on the extent to which it promotes or hinders satisfaction of (D).

But now even with a deflationist approach to truth, we have been able to identify a value that intellectual virtues exist to promote. We have put the virtues in the position of handmaidens to enable us to satisfy instances of the schema (D). In other words, they are handmaidens to the truth. And, by the standard of the debate in ethics, this is to throw in the towel on behalf of anything worth calling virtue epistemology.

4. Virtue and Reason

There is, I think, only one way in which these conclusions could be resisted. The priority of truth in the assessment of traits as virtues would need to be admitted, at the level at which we have been considering it. But, it would be maintained, this is only superficial. At a *deeper* level, it is the virtues that *give* us our conception of truth. This is, in effect, to return to the priority that we initially criticized in Peirce. Truth itself will be understood in terms of the upshot of virtuous inquiry.

But we have already said, on behalf of minimalism, that truth is not to be understood in any such grand way: The involvement of truth only came as a way of generalizing the desirability of satisfying the (D) schema. Individual instances of this schema gave individual goods; talk of the value of truth merely serves as a way of summing them up.

So the suggestion has to be that something Peircean is *concealed* within (D), or within its instances. And there is one obvious place to look, which is where the idealist tradition has always looked, namely, at the nature of judgment itself. (D) takes

the proposition that p for granted. But suppose, as seems plausible, that propositions are a kind of abstraction from the nature of judgment, and that judgment is an activity somehow constituted by what counts as exercising virtue in doing it, just as chess is an activity defined by what counts as winning. Then, even given minimalism, we have the necessary set of priorities: Virtues give us judgments which give us truth.

There are individual spheres in which this kind of suggestion may work especially well. Color perception, and secondary qualities in general, are perhaps the favorite. Here it is plausible to suggest that truth is somehow constituted by good practice in judgment. The variety of "response-dependent" analyses on the market give ways of filling out this thought. The truth that there is a smell in the room is not something further or over and above the truth that good receptors find it smelly. The truth that a surface is red is not a distinct fact from the fact that good practice in the way of color-judgment certifies it as red. Here practice is identified as good in terms of virtues: close attention, restriction to a privileged kind of light, ability to generate consensus, and so on (there are fewer marks of good practice in the case of smells). Anyone essaying a color "judgment," but who did not realize that the correctness or incorrectness of his verdict was hostage to satisfying desiderata such as these, would be convicted of being not part of the practice, not really, therefore succeeding in making the judgment at all.

Notice that this kind of thought does not stand in the way of the relevant example of schema (D):

(D_c) It is good that, if the lights are red then I believe that the lights are red; and it is good that if I believe that the lights are red, then the lights are red.

For there is sufficient distance between the lights being red and my believing it for there to be a chance of these conditionals failing, and it is better if they do not. But that is consistent with the truth that the lights are red being constituted by the fact that best judgment would determine them as red. And here "best" can be filled out by the other virtues that govern the practice: sustained, repeatable, consensus-generating, and so on. The gap that closes, but rightly on this approach, is any between what best practice would have us believe and the truth.

Should we generalize the secondary-quality case? Some philosophers believe that all concepts are "response dependent" in the way that colors plausibly are.[7] But perhaps we can think of the color case as illustrative, while preserving some difference between primary (or tertiary) qualities and concepts and secondary qualities and concepts. For it is not clear that the assertibility conditions in question need always to mention our responses. We could talk of the circumstances in which the virtuous are warranted in confidence, without thinking that it is the responses of the virtuous that in any way constitute the concept or property in question. So, for instance, the proposition that a shape is circular has its verification conditions. The virtuous only make such a judgment in the light of successfully completing or contemplating the completion of quite determinate procedures. But it need not follow that the responses of the virtuous themselves "constitute" the shape in the way that,

arguably, the responses of the virtuous constitute the fact that something is red. Clearly there is much more to be chewed over here: For the moment I am only interested in gesturing at what seems to be a possible theoretical space rather than arguing that we ought to inhabit it.

The standard way of generalizing would be to identify truth with warranted assertibility. But the difficulties of that proposal are formidable and well canvassed. Philosophers who have at various times promoted "warranted assertibility" accounts of truth have tended to diminished enthusiasm as time goes on.[8] I think a more plausible line would be to accept the gift offered by minimalism and refuse to work in terms of any kind of reduction or analysis of truth itself. Instead, *propositions* or *judgments* would be located in terms of their evidential relations. This is, I think, the way Horwich himself counsels us to look at it. In his terms, we locate judgments first by a use theory of meaning, and second by identifying the fundamental feature of use, the "basic acceptance property" that governs a speaker's overall use of the terms involved in making the judgment.[9]

The picture that this presents is quite in line with a strong virtue epistemology. Each judgment (or perhaps constituent of a judgment) has its own conditions of acceptance. Epistemic virtue will require conforming your own disposition to accept the judgment to those conditions for acceptance. In other words, anyone essaying a judgment is in a space of acceptance conditions that will dictate norms for proper acceptance, and hence the virtue or vice involved in accepting the judgment in particular circumstances. But Putnam is also right in supposing that this development of minimalism carries verificationist costs; whether those costs are bearable is clearly too big a question to settle now.[10]

Horwich himself believes that use is a purely factual concept, albeit one that has normative implications. That is, the "basic acceptance properties" attach to terms because of the use we actually make of them, not any more idealized concept of what the virtuous use of them requires. Nevertheless, just as other facts have normative implications (not entailments) so does this kind of fact. I think for present purposes we also do not have to solve whether Horwich is right about this, or whether normativity is more integrated with concepts of meaning and reference than he allows. Either way, to make a judgment is to be susceptible to criticism as epistemically virtuous or not. And, it seems, a strong epistemic virtue theory can be defended on this picture. For truth is no longer standing as an external, independent goal to which virtue tries to conform. Rather, each judgment comes with its own "virtuous acceptance conditions": the basic conditions governing what situations allow for proper confidence in it.

In "The Folly of Trying to Define Truth," Davidson attacked minimalism not on the grounds that a definition of truth is available, but on the grounds that there is a circle of terms such as judgment, proposition, truth condition, of which we need *some* philosophical account.[11] This is indeed the position that I arrived at in *Spreading the Word*.[12] But the Horwich of *Meaning* is not, it seems to me, a proper target of Davidson's attack. For here there is a philosophical account of meaning, and it purports to enter the entire circle of meaning terms as a whole, just as Davidson (and Putnam) suggest.[13] This is, so far as I can see, the only way in which a strong virtue

epistemology could be pursued. And we might suggest that seeing it like this will deliver one substantial benefit. The verification theory of meaning had an extremely limited view about what *kind* of virtuous acceptance conditions judgments could possess. They had to relate to experience in a particularly direct, stodgy kind of way, and that was all. Whereas with a more generous conception of what makes for virtuous acceptance, a more generous conception of meaning and meaningfulness opens up. There is no obstacle to bringing in virtues of reason and of theory, reinstating, perhaps, a priori propositions and certainly theoretical propositions. And this must be counted a substantial gain. Insofar as virtues are heterogeneous and subtle, so can judgments be.

5. Epistemology with an Attitude

So far I have said little about what it is to deem a cognitive trait a virtue. Our concern has been simply to explore the relationship between so deeming it, on the one hand, and thinking of it as conducive to the maximization of truth, on the other. I have concluded, although tentatively, that we need not suppose that this relationship undermines moderately strong epistemic virtue theory. But the ingredients we had to bring on board to secure that result may not appeal to everyone. They include minimalism about truth, and a "use theory" of meaning in a version in which use is primarily identified by virtuous verification or assertibility conditions. And these are sinister allies. Furthermore, even they do not enable us to reverse the priority of truth over virtue, as the strong program demands. At least in the Davidsonian form, they can at best give us the "virtuous circle" account, whereby judgment, truth, and epistemic virtue come as in a rush. This may be progress, at least compared to very weak virtue epistemology, but it is not a vindication of the strong program.

However the chips fall here, there is a question of what else is involved in deeming a cognitive trait to be a virtue. Here, I would argue, the way is a little clearer. To deem a trait a virtue is interchangeable with deeming some situations to be ones in which a judgment is certain or reasonable. The virtuous person is simply the person who discriminates such situations rightly and forms his confidence in conformity with them. So the question is in effect identical with that of what it is to discriminate what reason requires in different circumstances.

Profiting from work in ethics, we can see it like this. Being responsive to reasons means adjusting confidence in the *right* way, just as acting on reason means adjusting action in the right way. Distinguishing a circumstance as one that *calls for* such an adjustment is a matter of privileging it. And this is a matter of practical attitude. It is a matter of endorsing one kind of movement of the mind, or of ruling out other putative movements of the mind. This endorsement can come in different degrees, from something rather weak, like regarding an inference as permissible but not obligatory, up to something very strong, like regarding an inference as obligatory and dissent as crazy.

When we talk of our reasons either for believing or acting, we could be simply reporting on the causal background for our ending up as we are. But normally we

are doing more: we are in a normative space rather than a purely descriptive space. This means endorsing what has moved us as the kind of thing that can permissibly or obligatorily move people. But of course selecting something for that privilege is itself something that we do. It is expressing an aspect of our stances toward intellectual or practical movement.

Some philosophers cannot comprehend this. They want to keep the "ought" of reason free from contamination by the natural world.[14] But there is no contamination, and nothing supernatural needed to fend it off. As Kantians are fond of pointing out, we can indeed notice that something moves us and then stand back and ask whether it ought to be moving us as it does. The "is" does not settle the "ought." But is's settle what we will take the oughts to be. That is, when we do end up privileging one movement, and endorsing it as reasonable and either permissible or obligatory, this will be a matter of our own psychologies: of the movements with which we can feel comfortable. And of course, at the bottom of things we may be sadly aware that comfort is about all we have. That is, suppose the last word about induction is that it is just custom and habit, or the last word about theory is that it strikes us (now) as compulsory, or the last word about the a priori is that we cannot imagine it otherwise. Then we will be left realizing that our powers of critical reflection are at best limited compared to the grip of natural habits. We might fantasize about standing at a greater distance from ourselves, but here, perhaps even more than in the case of ethics, we are condemned not to do so.

Notes

1. My impression is that many writers do not notice or care about the difference. Zagzebski quotes Kvanvig, Armstrong, Nozick, Goldman, and Sosa as writers who have, in effect, identified virtue epistemology with reliabilism. See her *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996): 10–11.

2. The latest attempt to hold this ship together is that of Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999): chapters 9 and 10.

3. John McDowell, *Mind and World* (Cambridge, Mass.: Harvard University Press, 1996): 8.

4. I say modern critics because it is not at all plain to me that this was Sellars's own problem. He seemed more concerned to argue against the idea that any given, qualia or not, could be the basis of an infallible inference to a piece of propositional knowledge that is thereby rendered incorrigible.

5. This is the formulation given by Paul Horwich, *Truth* (Oxford: Oxford University Press, 1999).

6. Stephen Stich, *Deconstructing the Mind* (New York: Oxford University Press, 1996).

7. Philip Pettit, *The Common Mind* (Oxford: Oxford University Press, 1993), espouses a version of this Coming from quite different considerations, so does Jerry Fodor, *Concepts: Where Cognitive Science Went Wrong* (Oxford: Oxford University Press, 1998): 137. For criticism of Fodor's position, see Fiona Cowie, *What's Within?* (New York: Oxford University Press, 1999): 93–111.

8. I am thinking especially of Hilary Putnam Michael Dummett never came out wholeheartedly in favor of such theories, although he clearly recognized them as immensely attractive.

9. Paul Horwich, *Meaning* (Oxford: Oxford University Press, 1999): 44.

10. Hilary Putnam, *The Threefold Cord: Mind, Body and World* (New York: Columbia University Press, 1999): 51.

11. Donald Davidson, "The Folly of Trying to Define Truth," *Journal of Philosophy* 93 (1996): 263–78.

12. Simon Blackburn, *Spreading the Word* (Oxford: Oxford University Press, 1984): 264–81.

13. Putnam, *The Threefold Cord*, 70.

14. I should say that Tom Nagel, Barry Stroud, Chris Korsgaard, and Jean Hampton are each afflicted with this dualism, but it is really the common property of Aristotelians and Kantians.

3

THE UNITY OF THE
EPISTEMIC VIRTUES

Alvin I. Goldman

1. Unitarianism and Its Rivals

One of the central questions about the moral virtues that preoccupied Socrates concerned the unity of the virtues. Are the several virtues aspects of a single virtue or are they entirely distinct and independent? Socrates himself apparently accepted the doctrine of the unity of the virtues. In this essay I shall explore the tenability of an analogous doctrine for the intellectual, or epistemic, virtues, namely, that the various epistemic virtues are all variations on, or permutations of, a single theme or motif. I am not confident I can make this doctrine stick in full and complete detail, but I want to explore how close one can come in defending its plausibility.

What does one mean by an epistemic virtue, and what kind of unity might one hope to find in this territory? Almost everyone agrees that a virtue is an excellence, but which types of things count as excellences? In the Aristotelian tradition, an excellence is some kind of ability, disposition, power, faculty, or habit. Epistemologists like Ernest Sosa and John Greco seem to adopt a heavily Aristotelian conception of excellences in their epistemological uses of virtue theory. They characterize intellectual virtues as mental faculties, powers, or abilities to produce beliefs that are true (Sosa 1985, 1988, 1991; Greco 1992). In my own previous writing from a virtue perspective, I slightly expanded the possible scope of virtues by including mental *processes* (process types) as well as faculties, powers, or competences (Goldman 1992a). In the present essay I wish to be more inclusive yet and allow even types of *action* to be subsumed under the epistemic virtues. This would accord with John Locke's usage, who talks of "virtue" and "vice" as applying to actions: "[M]en everywhere . . . give the name of virtue to those actions, which amongst them are judged praiseworthy; and call that vice, which they account blameable . . ." (1975, II: 28).

What kind of unity might we hope to find in the sphere of epistemic virtues? Two strong versions of virtue unitarianism would be (1) an identity theory and (2) an inseparability theory. An identity theory is what Socrates endorsed, the view that all the (moral) virtues are really one and the same.[1] I would not dream of endorsing

so strong a theory for the epistemic virtues. Nor am I tempted by an inseparability theory, which would claim that you cannot possess one epistemic virtue without possessing the others. The type of virtue unitarianism I wish to explore is the more modest claim of *thematic* unity: the various epistemic virtues all share, or are derived from, some common unifying theme (or themes). A moderate form of virtue unitarianism would say that all virtues stand in a single relation, for example, a causal relation, to a common value, goal, or desideratum such as true belief. This is roughly the view shared by Sosa, Greco, and myself in earlier writings. A still weaker form of virtue unitarianism would not demand a single relation to a single common end. First, it would permit a plurality of relations to the specified value or desideratum; and second, it might even tolerate a smallish number of intimately related values rather than a single value. (We could not allow too many values, or entirely diverse values, without letting unitarianism collapse into pluralism.) The form of virtue unitarianism I shall try to defend in this essay is somewhere between a moderate and weak form of unitarianism. The principal relation that epistemic virtues bear to the core epistemic value will be a teleological or consequentialist one. A process, trait, or action is an epistemic virtue to the extent that it tends to produce, generate, or promote (roughly) true belief. But this causal relation is not the only one that will figure in my story.

Some proponents of "high church" virtue epistemology might find elements of teleology or consequentialism anathema to their hopes for a distinctive, virtue-based epistemology. By "high church" virtue epistemology, I mean a form of virtue epistemology that models itself closely after virtue ethics, which many theorists view as a rival to ethical consequentialism and deontologism. I think we should resist the temptation to insist that virtue epistemology must conform to the model of ethical theory; in this I depart from some virtue epistemologists such as Linda Zagzebski (1996). Epistemology and ethics are different fields, and it should not be presumed that what holds in one must also hold in the other. Nor is it clear, for that matter, that virtue ethics must eschew consequentialism and deontologism (see Hursthouse 1996, Trianosky 1990).

In the remainder of the essay, I shall attempt to make a case for the unity of epistemic virtues in which the cardinal value, or underlying motif, is something like true, or accurate, belief. I call this view *veritism* (see Goldman 1999a). This position has two types of rivals. The first is pluralism, which denies any thoroughgoing unity among all the epistemic virtues. The second type of rival is any theory that champions an alternative unifying theme, different from truth. Veritistic unitarianism must fend off challenges from both directions.

2. Veritistic versus Justificational Value

An obvious challenge to veritistic unitarianism arises from the fact that, on everyone's theory, justified belief is a distinct state of affairs from true belief but a preeminent example of an epistemically meritorious or valuable state of affairs. Unless the admittedly distinct state of justified belief can be shown to have its value in some

derivative fashion from the value of true belief, veritistic unitarianism is in trouble. If justificational status has to be posited as a value entirely autonomous and independent of truth, it looks like pluralism wins the day. Or perhaps justifiedness could even replace true belief as the core epistemic value.

The obvious strategy for veritistic unitarianism is to defend a reliabilist theory of justification, or at least some form of truth-linked justification theory. The idea would be that true belief is the ultimate value in the epistemic sphere, and various belief-forming processes, faculties, or mechanisms are licensed as virtuous because they are conducive to true belief. Beliefs are regarded as justified when they are produced by these very truth-conducive processes (or processes thought to be truth-conducive), even on those occasions when the beliefs are false. In other words, justified belief is a separate and independent value from true belief; but beliefs qualify as justified precisely because their provenance is that of truth-promoting processes. So their ultimate source of value remains veritistic.

There are, of course, many alternatives to reliabilism. In the rest of this section I shall examine the chief traditional alternatives: foundationalism and coherentism. There is no space here to try to prove that any viable form of foundationalism or coherentism *must* invoke veritistic value. That would be a book-length project. Instead, I shall mainly content myself with the limited observation that many prominent forms of these approaches *do* invoke veritism. In other words, I shall show how these other approaches are also, quite frequently, steeped in the waters of veritistic value.

Starting with foundationalism, we first note that many foundationalists pledge allegiance to true belief as the prime epistemic desideratum. Chisholm says that it is one's intellectual duty to try his best to bring it about that, for every proposition he considers, he accepts it if and only if it is true (Chisholm 1977: 14). It is also common for foundationalism to provide conditions for basic justification that are linked to truth. In the case of infallibilist foundationalism, a belief is basic if it has some characteristic that necessarily guarantees its truth. In the case of fallibilist foundationalism, beliefs can qualify as basic without necessarily guaranteeing truth, but the status of basicness may nonetheless be conferred by some explicit or implicit form of truth-indicativeness. Beliefs about introspectively accessible subject-matter — one's own current pains or other mental states — may qualify as basic because introspection is highly reliable. Perceptual beliefs may be deemed basic because perceptual appearance states are usually indicative of environmental truths. To be sure, there are accounts of basicness that appeal to such notions as "self-justification," which make no reference to truth. But it is questionable whether a plausible account of self-justification can be given that doesn't tacitly invoke truth-indicativeness. My suggestion, then, is that foundationalism's best prospects for success rest on an appeal to true belief as the fundamental epistemic value in terms of which its distinctive notion of basicness must be defined.

I turn next to coherentism. Some coherentist theories wear their underlying veritism on their sleeves. BonJour's (1985) defense of coherentism makes no attempt to hide the fact that his rationale for a coherence criterion of justifiedness is that coherence is a reliable indicator of truth. BonJour writes:

> The basic role of justification is that of a *means* to truth. . . . If epistemic
> justification were not conducive to truth. . ., if finding epistemically justified be-
> liefs did not substantially increase the likelihood of finding true ones, then epi-
> stemic justification would be irrelevant to our main cognitive goal and of dubious
> worth. It is only if we have some reason for thinking that epistemic justification
> constitutes a path to truth that we as cognitive beings have any motive for prefer-
> ring epistemically justified belief to epistemically unjustified ones. (1985: 7–8)

It would be hard to find a more explicit endorsement of veritism.

Lehrer's form of coherentism also has a rich strain of veritism in it, although it takes a bit of probing to see all of the veritistic themes. First, Lehrer explicitly endorses true belief (or acceptance) as the cardinal epistemic aim involved in justification. He characterizes the "objective of justification" as "accepting something if and only if it is true" (Lehrer 1990: 82). Second, the fundamental doxastic concept in his theory of justification is the concept of "acceptance," defined as a propositional attitude that arises from the purpose of obtaining truth and avoiding error (Lehrer 1974, 1989, 1990). Third, truth is salient in Lehrer's account of "verific justification," which he defines in terms of an acceptance system that is obtained by deleting statements of the form, S accepts that p, when p is false.[2] Fourth, a pivotal role in his theory of justification is played by an ultimate first principle "T", namely, "I am a trustworthy evaluator of truth." Lehrer writes: "What I mean by saying that a person is a trustworthy evaluator of truth and error is that when she accepts something as true . . ., her accepting what she does is a trustworthy guide to truth in the matter" (1989: 143). The fact that it is critical to one's being justified that one be a trustworthy guide to truth, or at least that one be justified in believing that one is a trustworthy guide, signals that the fundamental goal of the enterprise is precisely to accept the truth, or at least avoid error. So veritism really pervades Lehrer's theory. In fact, in one place he calls himself a sort of reliabilist, just not a *causal* reliabilist. "I agree with [Goldman] that reliability or probability is central. But it is the state of accepting something that must be a reliable or trustworthy guide to truth rather than the process that originates or sustains acceptance" (1989: 147).

Are there varieties of coherentism that offer decidedly non-veritistic values? There are certainly authors whose endorsements of coherentism make no explicit appeal to truth or error avoidance. The question is whether their theories provide a compelling account of justification if they are sharply disconnected from truth. For example, Gilbert Harman's brand of explanatory coherentism contends that all inductive inference is inference to a total explanatory account. "Induction is an attempt to increase the explanatory coherence of our view, making it more complete, less ad hoc, more plausible" (1973: 159). Such an inference is warranted, Harman intimates, if the resulting total view possesses more explanatory coherence than competing total views would have. Although Harman does not use the following terminology, he may be taken to imply that explanation and coherence are the epistemic *values* we seek to maximize, not true belief or error avoidance.

But let us press more deeply. Why do we suppose—granting for the moment that we do—that inferring the view with the greatest explanatory coherence is war-

ranted? I suspect it is because we assume that views that maximize explanatory coherence are most likely to be *true*. If this is correct, then the ultimate goal or value associated with warrant is that of true belief. Isn't this indeed suggested by some of Harman's own language? Isn't a "less ad hoc" view more likely to be true? Isn't a "more plausible" view more likely to be true? So it is far from clear that this analysis steers us in a direction away from veritism.

3. Evidence Proportionalism and Evidence Gathering Virtues

Another possible approach to the theory of justification—which need not be committed to either foundationalism or coherentism—is what I shall call *deontological evidentialism*. Deontological evidentialism, as I conceive of it, says simply that an agent should assign a degree of belief to a proposition in proportion to the weight of evidence she possesses. If the weight of her evidence is strong, her degree of belief should be substantial; if the weight of her evidence is weak, her degree of conviction should be proportionally muted. A proponent of this approach would hold that the requirement of proportioning is a purely deontological one, not derived from any consequentialist consideration, such as the thesis that proportioning leads to truth. I am not certain which current philosophers, if any, endorse deontological evidentialism. It might be the position of Richard Feldman and Earl Conee (1985), and perhaps Richard Jeffrey's (1992) "radical probabilism" is a species of it.[3] A crucial feature of the approach, for present purposes, is that it would constitute a rival to veritism. It would not rationalize proportionment as a means to true belief, error avoidance, or any other further end, but would treat it as an independent principle of "fittingness."

The main problem facing deontological evidentialism is to account for the virtues of evidence gathering. If proportioning your degree of belief to the weight of your evidence is the sole basis of epistemic virtue, cognitive agents can exemplify all virtues without gathering any evidence at all by working with the most minimal quantities of evidence. According to deontological evidentialism, it is just as meritorious for an agent to adopt a doxastic attitude of "suspension" when her evidence is indecisive as to adopt a doxastic attitude of full conviction when her evidence is quite dispositive. Both are equally good instances of proportioning degree of belief to the weight of one's evidence. No further epistemic merit or praise can be earned by investigation, research, or clever experimentation, the outcome of which might discriminate between competing hypotheses. In short, deontological evidentialism is perfectly content with investigational sloth! This is surely a major weakness in the theory, because numerous epistemic virtues are to be found among processes of investigation. When a scientist performs a clever experiment that selects among otherwise equally plausible hypotheses, she earns some of her profession's strongest (epistemic) kudos. Good experimental design is at the heart of scientific, and hence epistemic, progress. Deontological evidentialism has no way to accommodate virtues like clever experimentation. Deontological evidentialism implies that a scientific

community's epistemic position is just as satisfactory in the absence of revealing experiments as with them, because scientists can accurately proportion their degrees of belief to their evidence in either case. Without experiments, they can suspend judgment; with experiments, they can favor experimentally supported hypotheses. In this way, appropriate proportionment is attainable.

Veritism's resources for handling this problem are obviously more promising. The virtue of well-chosen observation or experimentation lies in its production of experience that tilts in favor of one hypothesis (or perhaps a family of hypotheses) over others. A cognitive agent who appreciates this evidence can raise her degree of belief in the favored hypothesis in response to this evidence; perhaps she will actually accept it. In a favorable case, when the accepted hypothesis is true, she will achieve an outcome—true belief—that is the mark of value under veritism. In general, virtues of good investigation are among the cardinal intellectual virtues. Veritism neatly accounts for such virtues, whereas deontological evidentialism draws a blank in trying to rationalize their intellectual merit.

It may be argued that deontological evidentialism is correct as a theory of *justification* but a separate theory is needed as an account of proper *inquiry*. A distinction between these two subject matters is emphasized by Susan Haack (1993). However, Haack's own conception of the norms of inquiry is distinctly veritistic. She writes: "The goal of inquiry is substantial, significant, illuminating truth" (1993: 203). So her approach does not obviously differ with what I have said above, in terms of its fundamental rationale for the conduct of inquiry.

However, completely different types of rationales for new evidence gathering can be found in the Bayesian literature. To address this question, I. J. Good (1983, chap. 17) produced a proof that making new observations will maximize expected utility. "*In expectation*, it pays to take into account further evidence, provided that the cost of collecting and using this evidence . . . can be ignored" (1983: 178, emphasis added). But why should inquiry be directed at increasing *expected* utility? Expected utility is a function of subjective probabilities, and these can be as wild or misguided as you please. As an *aim* of inquiry, expected utility is not well chosen.

This is not to say that new evidence gathering coupled with Bayesian reasoning has no suitable rationale. In fact, if we return to the veritistic perspective, it can be shown that Bayesian reasoning from new experiments yields an *objectively expected* increase in truth-possession—when certain additional conditions are met. Moshe Shaked and I have proved that if an inquirer has conditional probabilities (likelihoods) vis-à-vis a prospective experiment that meet certain conditions, there will be an objectively expected increase in his degree of truth-possession if he performs this experiment and reasons from its observed results in a Bayesian fashion (Goldman and Shaked 1993).[4] This again points to the fact that the most promising way to rationalize the intellectual virtue of designing and performing good experiments lies in the veritistic direction.

I conclude that veritism is the best way to account for the obvious fact that among our chief intellectual virtues—certainly our chief *scientific* virtues—are the clever design and execution of observational or experimental procedures.

4. Monistic versus Dualistic Veritism

I have been touting veritism as a promising unitarian approach, but is it really unitarian? Haven't we been brushing under the rug the fact that veritism really posits, not one, but *two* epistemic values: both true belief and error avoidance. These are distinct values, not reducible to one another, so how can I claim that this pair of values promises to realize the unitarian thesis?

This criticism was already anticipated when I outlined different possible strengths for a unitarian doctrine. Weak unitarianism, I indicated, is a brand of unitarianism that can invoke more than one fundamental epistemic value, as long as the several values have an intuitive homogeneity or integration. That certainly seems to hold for the two values of true belief and error avoidance. However, it is not yet clear that veritism must resign itself to what might be called "dualistic" unitarianism. There is a way to blend the two traditional veritistic values into a single *magnitude* or *quantity* of veritistic value, yielding monistic unitarianism. I have attempted this in *Knowledge in a Social World* (Goldman 1999a) and shall further defend it here.

Before turning to the categorical, or binary, concept of belief, let us consider degrees of belief, scaled on the unit interval. Degree of belief 1.0 is the highest degree of belief: subjective certainty. Degree of belief 0 is the highest degree of *dis*belief. This level of disbelief in proposition p is also, equivalently, the highest level of belief in not-p. Degrees of belief at or near 0.50 represent a maximum of subjective *un*certainty; they represent suspension of judgment, or the absence of an opinion. Now, when we try to move from degrees of belief to belief *simpliciter*, it is unclear what the threshold should be. Perhaps any degree of belief in p above 0.80 qualifies as "belief"; or perhaps it is 0.90. It will not matter where, exactly, we set the threshold, but whichever number x we choose as the threshold for belief, we should choose $1.0 - x$ as the threshold for disbelief. If 0.80 and above counts as belief, then 0.20 and below counts as disbelief, or belief in the negation of the proposition.

Let me now turn to values associated with different degrees of belief in true or false propositions. I propose that the highest degree of belief in a *true* proposition counts as the highest degree of "veritistic value" (with respect to the question at hand, e.g., whether p or not-p is the case). In general, a higher degree of belief in a truth counts as more veritistically valuable than a lower degree of belief in that truth (see Goldman 1999a, pp. 88–89).

This conception of veritistic value is readily applied to the more traditional but coarse-grained categories of belief, suspension of judgment, and disbelief as special cases. Whatever the exact threshold for belief may be (greater than 0.50), believing a truth carries more veritistic value than suspension of judgment; and suspension of judgment carries more veritistic value than disbelief. Now, disbelieving proposition p is equivalent to believing proposition not-p. And when p is true, not-p is false. So suspending judgment vis-a-vis a true proposition p has more veritistic value than believing the false proposition not-p. Thus, the intuitive rank-ordering of veritistic value is confirmed: True belief is preferable to suspension of judgment, which is preferable to false belief (error).

Various measures of truth-related value seem to presuppose something along the

foregoing lines, although these measures don't always exploit all features or details of the foregoing scheme. For example, reliability is a truth-linked measure that gives positive weight to true belief and negative weight to false belief, but it ignores suspensions of judgment. The "power" measure I have proposed elsewhere (Goldman 1986, chap. 6) properly reflects the inferiority of suspension of judgment to true belief, but it wrongly treats suspension of judgment and error as essentially equivalent. Thus, the veritistic value scale proposed above seems preferable, and it has the welcome feature of presenting a *single* magnitude that can serve as the underpinning for a pleasingly *monistic* version of veritistic unitarianism.

Does this scheme of veritistic value accord with commonsense notions about intellectual attainments? I think it does. If a person regularly has a high level of belief in the true propositions she considers or takes an interest in, then she qualifies as "well-informed." Someone with intermediate levels of belief on many such questions, amounting to "no opinion," qualifies as uninformed, or ignorant. And someone who has very low levels of belief for true propositions—or, equivalently, high levels of belief for false propositions—is seriously misinformed. Since the terms "well-informed," "ignorant," and "misinformed" seem to reflect a natural ordering of intellectual attainment, our scheme of veritistic value seems to be on the right track. I think we would also find that many ordinary expressions that designate intellectual virtues refer to processes or traits that promote well-informedness, whereas expressions that designate intellectual vices refer to traits that promote ignorance or error.

5. "Pragmatic" Virtues, Science, and Interest-Responsiveness

Against the veritistic picture I have been painting, there is a panoply of competing cognitive virtues often touted by epistemologists and especially philosophers of science. The latter also offer special reasons for doubting the role of truth in our system of epistemic values. These challenges must now be confronted.

W. V. Quine and Joseph Ullian (1970, chap. 5) list five "virtues" of a hypothesis that count toward the plausibility of a hypothesis. In different terminology, these might be considered "values" for appraising cognitive practices. I'll concentrate on the first three virtues. The first value is *conservatism*. One scientific hypothesis is preferred to another, they say, if it requires scientists to abandon fewer of their previous beliefs. Their second value is *generality*, illustrated by Newton's theory of gravitation. The ability of a hypothesis to explain a wide range and/or variety of phenomena makes it specially worthy of our credence. Their third value is *simplicity*, a widely invoked cognitive value in the philosophy of science literature. Many other epistemologists and philosophers of science echo these themes. In a variant formulation of conservatism, Harman enunciates a principle which says that a person is justified in continuing to accept a prior belief if he lacks a special reason to disbelieve it (Harman 1986: 46). In themes related to generality or explanatory power, we should mention Karl Popper (1962) and Isaac Levi (1980), who in their different

ways emphasize the contentfulness or informational value of a potential answer as an important cognitive desideratum. Similarly, Haack (as we have seen) and Philip Kitcher (1993) emphasize that it is not just truth that cognizers seek, but *significant* truth. Since the cited desiderata differ from truth or enlarge upon it, doesn't they undercut the prospects for epistemic unitarianism?

In response I begin by challenging conservatism, the notion that preserving prior beliefs is a worthy epistemic goal. I don't deny the descriptive thesis that cognitive agents prefer hypheses or theories that help them retain prior intellectual commitments. But this is simply because they are strongly *attached* to those commitments. They find it hard to be convinced that their old fabric of belief is error-filled, especially in radical ways. It is a philosophical mistake, however, to elevate this descriptive fact into an epistemic value. Furthermore, Harman's conservationist principle of justified belief does not capture any genuine component of the concept of epistemic justification. If someone acquires a belief at one moment by sheer guesswork or wishful thinking, that belief cannot suddenly attain justificational status in the next moment simply because the cognizer has no specific defeater for it.

I turn next to generality and simplicity. The claim I wish to make here, especially about generality, is that this is a *specialized* value peculiar to science, not one that pervades all cognitive inquiry. Science has a distinctive intellectual mission, which includes the attempt to find uniformities of nature and to explain as many phenomena as possible in terms of such uniformities. If more can be explained with less, so much the better. Perhaps the same distinctive mission accounts for the value of simplicity, though this is difficult to say since simplicity means different things to different theorists and no widely accepted analysis of simplicity has yet been found. The main point I wish to make is that epistemologists of science should not be allowed to persuade other epistemologists that distinctive goals and values of science are also goals of cognition and inquiry in general. When it comes to the prosaic purposes of everyday life, or specialties outside of theoretical science, there is no blanket premium on more general or comprehensive truths. We *sometimes* are interested in finding true generalizations, but finding such generalizations is not always more important to us than finding truths about particulars. This simply varies with our interests of the moment.

The moral to be drawn is that *interest* does play a role in evaluating cognitive practices and establishing cognitive virtues. When we are concerned with science, the distinctive interest of science in comprehensive theories therefore comes into play. But generality or simplicity should not be inserted into the pantheon of values for cognition in general. Admittedly, the dimension of interest does complicate our story. We can no longer suggest that higher degrees of truth-possession are all that count in matters of inquiry. But can't we incorporate the element of interest by a slight revision in our theory? Let us just say that the core epistemic value is a high degree of truth-possession *on topics of interest*.[5] Admittedly, this makes the core underlying value a somewhat "compound" or "complex" state of affairs. But, arguably, this is enough to preserve the idea of thematic unity and thereby preserve unitarianism. A further attractive feature of this approach is that it takes account of the fact that among the intellectual virtues are cognitive processes and practices that pro-

mote "interest-responsiveness." The ability to remain diligent, thorough, and persevere in addressing the questions that most interest us are important intellectual virtues (cf. Zagzebski 1996: 114).

Philosophers of science may not be so easily mollified by these maneuvers. There are bigger obstacles, many would contend, to a truth-centered account of epistemic virtue. For example, what can a veritist make of the fact that many so-called scientific "laws" are not true at all, and known not to be true; yet they continue to be invoked and used in science (Cartwright 1983)? Doesn't this demonstrate that science isn't so interested in truth after all? The same may be said of ordinary conversation and belief, where people are frequently imprecise and hence false in what they say and believe, yet this is blithely tolerated. Maybe *verisimilitude* (closeness to truth) can be substituted as a value in place of truth, but it is doubtful that any adequate concept of verisimilitude has yet been constructed.

My answer to this challenge is to distinguish the laws presented in textbooks and what scientists actually believe. At least in cases where it is known that the "laws" in question are only approximate, practitioners don't believe the contents of the stated laws. What is believed is not the lawlike formula L but instead something of the form, "Approximately, L", or "L holds for such-and-such ranges of application but not outside those ranges." Notice I am not saying that scientists ascribe the property of "approximate truth" to the laws in question. Rather, they have ways of mentally qualifying L to take account of its (known) imprecisions. What they believe (i.e., believe to be *true*) is some suitably hedged proposition. Concerning these hedged propositions, they do aim for truth, and want the contents they accept to be true. No doubt they are still wrong in many of these cases; but we are only talking here of *goals*, not *accomplishments*.

6. The Priority of Truth versus Justification

In this section and the next, I want to return to the matter of justifiedness. The defense of veritistic unitarianism has rested partly on the claim that the value of justifiedness is derivative from the value of truth. This claim might be challenged, however, by finding cases in which justifiedness takes precedence over truth. If such cases could be firmly established, wouldn't that deal a strong prima facie blow to veritistic unitarianism? How could a merely derivative value have priority over a more fundamental one? It would just prove that truth (true belief) isn't really more fundamental.

Before turning to cases, I want to question this style of critique. It is by no means clear that a derivative value could not assume some sort of priority over a more fundamental value. Suppose that a state with fundamental value is, for the most part, only *reachable*, *realizable*, or *accessible* to human agents via some action or state with derivative value. Moreover, the action or state with derivative value is more directly subject to "guidance" than the state with fundamental value. Then we might place greater weight on achieving or "performing" the more accessible action with merely

derivative value. We might even make such an action obligatory. This deontic force or obligatoriness would not necessarily indicate a greater *value* for the accessible action or state. It's just that we can't expect people to achieve the fundamentally valued state but we can expect them to take the best route in its direction. So we may positively *require* agents to take that route without also requiring them to achieve the more fundamental value. This would make it appear as if the required action or state is more important or weighty, and hence not merely derivative. But such an appearance would be deceptive.

Here is an illustration of this idea. In the legal arena jurors are required to vote for a given judgment—guilty vs. innocent, liable vs. non-liable—in light of the *evidence* they have heard at trial. They are not officially invited or required to vote in accordance with the *truth*. It might appear from this as if the legal system prizes evidence or justification above truth, and some might use this as a ground for saying that evidence or justification has priority over truth and is not merely derivative from it. But, as argued above, this would be a mistake. Truth *is* the main institutional goal of adjudication.[6] Since evidential justification is the only reliable route to truth, however, the system imposes an obligation on jurors to judge in accordance with the evidence. This institutional obligation does not negate the claim that truth is the fundamental value at which the adjudication arm of the law aims.

Readers prepared to concede this point for the legal realm might still be unpersuaded of it as a global thesis. Many might still say that the primary responsibility of an epistemic agent is conforming to justificational requirements, not getting the truth. These epistemologists would insist that justified belief, not true belief, is the primary epistemic value. Can this position be rebutted?

It may be illuminating to switch the focus of discussion from doxastic agency to other sectors of intellectual activity. Although traditional epistemology centers on credal acts, we should not forget activities involving speech or communication. When engaged in assertive discourse, people incur duties and responsibilities; they display virtues and vices. The principles governing such speech activities can shed light on the relationship between the twin desiderata of justification and truth.

Norms of assertive discourse are presented by Grice (1989) in the context of a theory of pragmatics; but Grice's norms can equally be seen as principles of social epistemology, which is how I treat them in *Knowledge in a Social World* (Goldman 1999a, chap. 5; also see Goldman 1994). Two of Grice's norms—he calls them "maxims"—are the following (see 1989, p. 27):

(1) Do not say what you believe to be false,

and

(2) Do not say that for which you lack adequate evidence.

What is the underlying rationale for these conversational norms? Grice postulates a cooperative venture of a quasi-contractual nature, where the venture's goals include the giving and receiving of information (1989: 29–30). This seems to me ex-

actly right and perfectly fits the veritistic approach, especially when it is understood that "information" entails *truth*. Conversation involves exchanges that can profit hearers because they can acquire true beliefs on the cheap, without being burdened by time-consuming or costly investigation. The posited desideratum—information or truth acquisition—makes excellent sense of the twin norms given above.

Norm (1) is straightforward under this approach. A speaker should not say what he believes to be false because what is so believed stands a serious chance of *being* false, and hearers will be at risk of becoming *mis*informed rather than informed. Norm (2) makes sense for the same reason, if we assume that adequate evidence is generally linked with truth. A proposition for which a speaker lacks adequate evidence stands a good chance of being false, even if the speaker believes it. So assertions unaccompanied by adequate evidence again put the hearer at risk of being misinformed rather than informed.

Can these maxims be rationalized under a different approach, one that gives exclusive or primary weight to justifiedness rather than truth? It might initially appear that this will work, certainly for norm (2) at any rate. How can a hearer acquire justified belief from a speaker's assertion unless the speaker himself has adequate evidence for what he asserts? On reflection, however, we can see that this is wrong. Hearers certainly can acquire justified belief in a proposition even if the speaker who asserts it neither believes it nor has adequate evidence for it. Suppose the hearer has every reason to trust the speaker; then the hearer is justified in believing what the speaker asserts. When a speaker knows that this is the hearer's circumstance, why should he obey either norm, even with the hearer's (epistemic) interest at heart? The hearer will still "profit" from the speaker's asserting something he neither believes nor is justified in believing; at least the hearer will profit in terms of the justificational desideratum, which is what the current approach emphasizes. Consider a second-grade school teacher who considers making various factual statements to his class for which they have no independent evidence one way or the other. The teacher believes all of these statements to be false, and has no justification for any of them. Nonetheless, since his pupils trust him completely, and have adequate reason for so trusting him, his making these statements will make the pupils justified in believing them. Why, then, shouldn't he go ahead and make those statements, according to the current—purely justificational—approach? There is, to be sure, a salient risk of their coming to believe falsehoods rather than truths, but that is irrelevant under the current approach.

A possible reply from the justification-centered camp is to introduce a *transmissional* conception of justification. According to such a conception, a speaker can make a hearer justified in believing a proposition only if the speaker himself *has* justification to *transmit*. Where the speaker lacks such justification, hearers cannot profit justificationally from his assertions. But is such a transmissional conception of justification an appropriate conception for someone who defends a justification-centered approach to epistemic value? I think not. A justification-centered approach is likely to appeal to internalists about justification. After all, externalist conceptions of justification standardly incorporate a link to truth and are therefore highly congenial to veritistic unitarianism. A justification-centered approach should appeal to in-

ternalists who wish to distance the concept of justification from the goal of truth. A transmissional conception of justification, however, is the antithesis of internalism. A hearer cannot tell if he gets justification from a speaker unless he can tell whether the speaker himself is justified, and that is a difficult matter to discern, something to which a hearer has no guaranteed access. Thus, a transmissional conception of justification is unlikely to suit a justification-centered theory; yet without that conception, such a theory cannot rationalize norms (1) and (2).

7. Varieties of Justification

The preceding discussion calls attention to the fact that, according to many writers, there is more than one sense or conception of epistemic justification.[7] If the prospect of multiple senses of justification is accepted, however, won't this undercut the thesis of veritistic unitarianism? Although some of these conceptions of justification—the more externalist ones, in particular—may comport nicely with veritistic unitarianism, other conceptions—especially the more internalist ones—might not comport with it at all. Indeed, the very existence of multiple justificational values seems to pose a challenge to unitarianism.

Now, I have already conceded that *monistic* unitarianism may not be defensible; so it wouldn't close the book on unitarianism if we had to concede more than one conception of justification. The more pressing question is whether all varieties of justification can be traced to the *veritistic* dimension of value. Purely externalist conceptions of justification—at least reliabilist versions of externalism—seem safe in this respect. But what about more internalist conceptions, or even mixed conceptions?

In my essay "Internalism Exposed" (Goldman 1999b), I give the example of Sally, who reads about the health benefits of broccoli in a questionable source, e.g., the *National Inquirer*. Despite this notorious source, Sally trusts the story and believes in the benefits of broccoli, though she encounters no further information that either corroborates or conflicts with it. At a subsequent time, Sally finds herself still believing in the health benefits of broccoli, but she no longer recalls where she learned about it. Is Sally's belief justified? I am inclined to say "no" (in this version of the case), because the only possible source of her justifiedness is a wholly unworthy source. Had she acquired her belief from the *New York Times* (another variant of the case I also consider), her belief would be justified. But in the *National Inquirer* version, I don't think the belief is justified. At any rate, it isn't justified in the "epistemizing" sense of justifiedness, in which justifiedness carries one a good distance toward knowledge.

Several readers of this example say that there is *another* sense of justifiedness in which Sally's belief *is* justified. Since Sally thinks she acquired the belief from a reliable source (though she doesn't recall what that source was), the belief seems to be justified *from her current perspective*. And that's all that matters for internalist justifiedness.

I don't want to dispute these readers' intuition that there is such a sense of justifiedness; the question is how to account for it. I offer an account that appeals to the

familiar *objective/subjective* distinction. Keeping matters simple, a belief has objective justifiedness if its genesis or sustainment is the result of appropriate processes (e.g., reliable processes). A belief has subjective justifiedness if, *by the agent's own lights*, the belief's genesis or sustainment is the result of appropriate processes. Sally's belief is not objectively justified but it is subjectively justified, because she takes her belief to have resulted from appropriate processes.[8]

The significance of this approach for veritistic unitarianism is as follows. Subjective justifiedness is a secondary concept, derived in an obvious way from objective justifiedness. This is the standard way of understanding subjective *X*ness for any (objective) concept of *X*ness (where *X* could be duty, rightness, etc.). Roughly speaking, the distinction is between how things *are* and how they *appear* from an agent's perspective. In the present case, a belief is objectively justified if it *does* meet certain standards that have a truth-indicative status. A belief is subjectively justified if it appears, from the agent's perspective at the time of belief, to meet those standards.[9]

If, as argued above, the standards for objective justifiedness are deeply rooted in the waters of truth-indicativeness, then subjective justifiedness, by implication, has a derivative element of truth-linkedness. This is not to say that being subjectively justified is a good indicator or means to truth. The relation to truth here is not a straightforwardly instrumental one. But I have not characterized unitarianism as the doctrine that all the concepts of epistemic value or virtue have the *same* relation to a single fundamental desideratum. The possibility of multiple relations was clearly anticipated (though a multiplicity of relations, it was conceded, would mark a weaker form of unitarianism). I am now proposing that subjective justifiedness exemplifies a distinct type of relation to the core veritistic desideratum, not an instrumental relation.

Another conception of internalist justification that differs from the purely subjective one is what I'll call the "higher-level" conception. One example is BonJour's insistence that justification must meet a "meta-justificational" requirement. *S*'s believing *p* is justified if and only if the belief not only meets some first-order standards of justification but, in addition, *S* is justified in believing that these standards are adequately truth-indicative (1985: 9). Now, I think there is a serious problem about how to formulate the pertinent higher-level condition. But if it embodies a second level of justifiedness, as the term "meta-justification" certainly suggests, and if justifiedness itself has a truth-indicative component, it certainly appears as if the higher-level approach to justification is strongly geared toward true belief as a pre-eminent goal. The relation between the higher-level condition and the truth desideratum may not be straightforwardly instrumental, but I have already argued in the discussion of subjective justifiedness that the relation to truth need not be instrumental to mesh with veritistic unitarianism.

8. Welfare Unitarianism

A completely different alternative to veritistic unitarianism than any considered thus far might center on some generic form of *moral* value rather than the distinctive

value of true belief. A familiar example of such a value is utility or welfare, but the list of relevant possibilities could be expanded to include justice, respect for rights, and so forth. To keep things simple, I shall explore this perspective by reference to welfare as the representative value. The idea would be that welfare not only underlies and unifies the moral virtues, but also the epistemic virtues. In fact, the epistemic virtues may not be distinguishable from the moral ones.

Some such thesis is suggested by Stephen Stich (1990, chap. 6) under the rubric of a "pragmatic" approach to cognitive evaluation. More precisely, Stich talks of cognitive value as residing in whatever people take to be intrinsically valuable. Since he rejects truth as the basis of cognitive evaluation, for a variety of reasons, he sees pragmatic values as a rival approach to veritism. Another example of welfare unitarianism is suggested by Philip Kitcher (1997), in discussing scientific decisions about whether to undertake certain kinds of research, especially research into the biological basis of human behavior. Kitcher presents two candidate decision rules for deciding whether to pursue such research (1997: 285):

(1) Pursue research if and only if the expected utility for the entire population is positive.
(2) Pursue research if and only if the expected utility for the underprivileged is positive.

These are obviously quite different, and Kitcher indicates his preference for (2). For our purposes the important thing to note is the contrast between *both* of these decision rules and veritism. Both rules invoke expected utility, not true belief, as the benchmark for scientific research decisions. To put it in terms of our present theme, utility or welfare would be the benchmark for determining a scientifically "virtuous" course of action.

None of these welfare approaches strikes me as a plausible approach to epistemic value or virtue. In particular, welfarism gives intuitively wrong results if it is applied to the question of *how* research should be conducted, as opposed to what topics should be the objects of research. There will almost certainly be cases in which conducting research in a certain way would pass a suitable welfare criterion though it would generate false beliefs; and conversely, there will almost certainly be cases in which research conducted in a truth-conducive fashion would not pass a welfare criterion. It seems wrong to say, however, that *epistemic* (intellectual) virtue would be exercised by adopting an anti-veritistic method of research in the service of optimizing the appropriate type of welfare.

To say this is not to deny that moral values might sometimes trump epistemic values, so that certain research should be sacrificed. Suppose a clever experiment has been devised that would probably yield extremely interesting data on human emotional responses, as all researchers in the area concede. Unfortunately, it would require infliction of severe pain on human subjects. Every scrupulous scientist will agree that the research should not be done; nor would it get past any human subjects committee. Is this because it lacks epistemic or scientific virtue, as the welfare approach would certainly say? No, there is a better explanation.

Although veritistic value is the fundamental benchmark of epistemic virtue, it is obviously not the *only value*. Nor is it the preeminent value for all purposes of life and action. Epistemological or scientific value sometimes conflicts with moral value, and when they conflict, epistemological value must give way. There is a moral "side-constraint" on scientific research, which is that the conduct of such research should not violate human rights or injure people. The human experiment described above should not be done, but not because it would lack scientific or epistemic value. Rather, it is a case in which the moral disvalue trumps the scientific value. Cases of moral trumping should not be an invitation to confuse scientific value with moral value; the truth desideratum should not be replaced with the welfare desideratum for the sphere of the epistemic. In rejecting welfare as the benchmark of epistemic value, I do not mean to reject considerations of welfare, or other measures of social significance, as one pertinent criterion in selecting *questions* or *topics* for scientific research. The role of "interest" has already been acknowledged in my own account of core epistemic value, and I do not exclude the notion that one factor that might determine a question's interest is the social significance its answer might have. The crucial thing, from my perspective, is that science is epistemically virtuous to the extent that it promotes a high degree of *truth* possession on questions of interest. What fixes appropriate interests (for science or other arenas of inquiry) is another matter.

The importance of this point is that veritism need not disagree with certain strands of a "value-ladenness" thesis in the philosophy of science, a thesis defended by feminist epistemologists such as Helen Longino (1990) and Elizabeth Anderson (1995). Veritism need not dispute at least one legitimate role in science for social and contextual values that is stressed by these writers, namely, the role of these values in placing questions on the scientific agenda. Anderson highlights the fact that scientists often judge the significance of questions in medicine, horticulture, and engineering in terms of practical social interests. Similarly, physicists investigate conditions for controlled and uncontrolled nuclear reactions and number theorists study algorithms that can rapidly factor very large numbers because of practical or contextual values (in the latter case the interest concerns the construction and decoding of encrypted messages). Veritism can completely agree with these points. It need not hold that science should always be driven by purely "internal" questions produced by its own puzzle-generating activities. It allows questions to be chosen for scientific research on the basis (at least in part) of practical social interests. Veritism will diverge from the value-ladenness perspective, however, if the latter suggests that the methods of "choosing" theories or fixing beliefs in theories (to answer the selected questions) may properly be guided by considerations that conflict with truth-based considerations. I am not certain what Anderson's or Longino's views are on this matter; but unless they also endorse this last-mentioned thesis, there is no necessary conflict with veritism.[10]

9. Conclusion

I would not claim to have made a thoroughly decisive case for epistemic unitarianism. Nonetheless, veritistic unitarianism has withstood quite a few challenges and

passed a number of tests with flying colors. While the specific form of unitarianism developed here is not the purest possible form, its tenability seems more defensible, on reflection, than many might have thought possible.[11]

Notes

1. Thus, T. H. Irwin writes: "Socrates believes that each virtue is identical to the very same knowledge of good and evil, and hence that all virtues are really just one virtue" (1992: 973–74).

2. See Lehrer (1990: 134–35, 149).

3. Susan Haack is an evidentialist but not a deontological evidentialist. She clearly offers truth-indicativeness as the basis for evidentialism: "The goal of inquiry is . . . truth; the concept of justification is specifically focused on security, on the likelihood of beliefs being true. Hence my claim that truth-indicative is what criteria of justification need to be to be good" (1993: 203). She defends her own specifically "foundherentist" approach to justification by saying that its criteria offer "the best hope for truth indication that we can have" (1993: 220–21).

4. Also see Goldman 1999a, chap. 4. The crucial condition is that the agent's subjective likelihoods must *match* the objective likelihoods. The result also assumes that the agent's priors are neither 0 nor 1.0, and the likelihood ratio is not 1.0 (the likelihoods are not identical). A second theorem is that *larger* expected increases in truth-possession are associated with greater divergences of the likelihood ratio from 1.0. In other words, the more "decisive" an experiment, the greater its objectively expected increase in truth-possession. What are here called "degrees of truth-possession" are discussed in the next section of the text under the label "degrees of veritistic value."

5. The role of interest is acknowledged in *Knowledge in a Social World*, sect. 3.5. The variety of potentially relevant interests discussed there makes for additional complications, but I don't think these are relevant to the present discussion.

6. For defense of this thesis, see Goldman 1999a, chap. 9.

7. A particularly strong version of this thesis is advanced by Alston (1993).

8. The most systematic development of a subjectivist account of justification (or rationality) is found in Foley (1987). In a later book, Foley acknowledges, in agreement with the present proposal, that 'justification' and its cognates "have both an egocentric and an objective side" (1993: 86). My essay "Strong and Weak Justification" (Goldman 1988) comes close to formulating the objective vs. subjective conceptions of justifiedness. But there I conceived of strong and weak justifiedness as "separate but equal," whereas here I want to suggest that subjective justifiedness is a secondary concept, derivative from that of objective justifiedness.

9. "Perspectivalism" is one conception of internalist justification that Alston (1989) considers.

10. Anderson does assign a role to contextual values in theory *choice*, but it is not so clear that the assigned role in any way cuts against truth. She says that contextual values come into play in telling us which *classifications* to use; but the choice of classifications does not obviously have a truth-impeding tendency.

11. Thanks to Todd Stewart for many valuable comments on initial drafts of the essay; to Julia Annas for instructive discussion of the virtue ethics literature; to Heather Battaly, who raised several useful points in her role as commentator at the Santa Barbara conference; and to Susan Haack, for commenting on (without necessarily agreeing with) my interpretation of her position.

References

Alston, W. (1989). "Internalism and Externalism in Epistemology," in *Epistemic Justification*. Ithaca: Cornell University Press.

———. (1993). "Epistemic Desiderata," *Philosophy and Phenomenological Research* 53: 527–51.

Anderson, E. (1995). "Knowledge, Human Interests, and Objectivity in Feminist Epistemology," *Philosophical Topics* 23.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Chisholm, R. (1977). *Theory of Knowledge*, 2nd ed. Englewood Cliffs, N.J.: Prentice Hall.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

Feldman, R. and Conee, E. (1985). "Evidentialism," *Philosophical Studies* 48: 15–34.

Foley, R. (1987). *The Theory of Epistemic Rationality*. Cambridge, Mass.: Harvard University Press.

———. (1993). *Working without a Net*. New York: Oxford University Press.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, Mass.: Harvard University Press.

———. (1988). "Strong and Weak Justification," in *Philosophical Perspectives* 2. Atasadero, Calif.: Ridgeview.

———. (1991). "An Economic Model of Scientific Activity and Truth Acquisition," *Philosophical Studies* 63: 31–55. Reprinted in Goldman (1992b).

———. (1992a). "Epistemic Folkways and Scientific Epistemology," in Goldman (1992b).

———. (1992b). *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, Mass.: MIT Press.

———. (1994). "Argumentation and Social Epistemology," *Journal of Philosophy* 91: 27–49.

———. (1999a). *Knowledge in a Social World*. Oxford: Oxford University Press.

———. (1999b). "Internalism Exposed," *Journal of Philosophy* 96: 271–93.

———, and Shaked (1993). "Commentary on the Scientific Status of Econometrics," in *Social Epistemology*, 7(3), 249–53.

Greco, J. (1992). "Virtue Epistemology," in J. Dancy and E. Sosa (eds.), *A Companion to Epistemology*. Oxford: Basil Blackwell.

Grice, H. (1989). *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.

Haack, S. (1993). *Evidence and Inquiry*. Oxford: Basil Blackwell.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

———. (1986). *Change in View*. Cambridge, Mass.: MIT Press.

Hursthouse, R. (1996). "Normative Virtue Ethics," in R. Crisp (ed.), *How Should One Live? Essays on the Virtues*. Oxford: Clarendon Press.

Irwin, T. H. (1992). "Plato," in L. Becker and C. Becker (eds.), *Encyclopedia of Ethics*, vol. 2. New York: Garland.

Jeffrey, R. (1992). *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Kitcher, P. (1993). *The Advancement of Science*. New York: Oxford University Press.

———. (1997). "An Argument about Free Inquiry," *Nous* 31: 279–306.

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

———. (1989). "Knowledge Reconsidered," in M. Clay and K. Lehrer (eds.), *Knowledge and Skepticism*. Boulder: Westview Press.

———. (1990). *Theory of Knowledge*. Boulder: Westview Press.

Levi, I. (1980). *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Locke, J. (1975). *An Essay Concerning Human Understanding*. P. Nidditch (ed.). Oxford: Clarendon Press.

Popper, K. (1962). *Conjectures and Refutations*. New York: Basic Books.

Quine, W. and Ullian, J. (1970). *The Web of Belief*. New York: Random House.

Sosa, E. (1985). "Knowledge and Intellectual Virtue," *Monist* 60: 226–45.

———. (1988). "Knowledge in Context, Scepticism in Doubt. The Virtue of Our Faculties," *Philosophical Perspectives* 2: 139–55.

———. (1991). "Reliabilism and Intellectual Virtue," in *Knowledge in Perspective*. Cambridge: Cambridge University Press.

Stich, S. (1990). *The Fragmentation of Reason*. Cambridge, Mass.: MIT Press.

Trianosky, G. (1990). "What Is Virtue Ethics All About?," *American Philosophical Quarterly* 27: 335–44.

Zagzebski, L. (1996). *Virtues of the Mind*. Cambridge: Cambridge University Press.

4

FOR THE LOVE OF TRUTH?

Ernest Sosa

Rational beings pursue and value truth (the true, along with the good and the beautiful). Intellectual conduct is to be judged, accordingly, by how well it aids our pursuit of that ideal. What does this mean, and is it true?

Even if intelligent life had never evolved or otherwise existed, Venus would still have orbited the Sun, so it would still have been true that Venus orbited the Sun. It is not the being thus true of what is true that we value indiscriminately. Some truths are good, but not all; far from it.

In loving the truth, then, what we value is not the being true of the truths. What we value in pursuing truth is rather our grasping it, our having it. What does this mean? Only through believing it does one relevantly have a truth: We have the truth that snow is white by believing that snow is white. In pursuing the truth what we want is (at least) true beliefs.

Suppose you enter your dentist's waiting room and find all the magazines missing. Deprived of reading matter, you're sure to doze off, but you need no sleep. Are you then rationally bound to reach for the telephone book in pursuit of truth? Were you not to do so, you would forfeit a chance to pluck some desired goods within easy reach.

If random telephone numbers do not elicit a wide enough yawn, consider a randomly selected cubic foot of the Sahara. Here is a trove of facts, of the form *grain x is so many millimeters in direction D from grain y*, than which few can be of less interest. Or take some bit of trivia known to me at the moment: that it was sunny in Rhode Island at noon on October 21, 1999. I confess that I will not rue my loss of this information, nor do I care either *that* or *how early* it will be gone.

As interpreted so far, the view that we rationally want truth as such reduces to absurdity, or is at best problematic.

What then is it we want in pursuing truth? If it is not after all true beliefs indiscriminately, what then is it? A manageable number of true beliefs? Obviously not; it is not just a matter of numbers. True beliefs that are not too costly? No, that also is refuted through our examples. True beliefs of a certain sort? What sort?

B

We are considering the role of the concept of truth in our intellectual lives and in epistemology. Does it have a role through one's motivation as a thinker?

What is the right model here? Is it the practical syllogism, with a truth-directed master practice? We focus on intellectual, belief-guiding *practices*; these may be social or individual, conscious or subconscious, freely formed or built-in or forced. What would be the master motivation required for proper guidance? What would be the content of the master practice? Would it be <If *P* is true, then one is to believe *P*>? No, the waiting room and Sahara examples show that we adhere to no such practice. Even if we did, moreover, it would seem pretty useless.

What other practice might use the concept of truth to guide our belief-formation? It might be suggested that we do after all prefer our beliefs to be true. We'd rather they were true than not true. Doesn't this show a generally shared pro-attitude favoring true belief?

You might reasonably want to repent your sins, to dine deliciously, to walk safely. Repented sins, delicious meals, safe walks — these are things one might well want, but in different ways. Sins you do not want; you just want to repent any you may commit. Meals you do want, properly spaced; and you would prefer your meals to be delicious. About walks per se your stance may be neutral. But if you have reason to take one, for exercise or pleasure or to get somewhere, you would prefer that it be safe. In each of these cases you want your *X*'s to be Ø, which is compatible with each of three attitudes on *X*'s per se: desire, aversion, and neutrality. The waiting room and Sahara examples suggest that our attitude on beliefs per se is neutrality. But this is compatible with our wanting our beliefs to be true, just as we want our walks to be safe.

The point must be put carefully. If we believe that a dear friend is terminally ill we would *not* want our belief to be true. What we want, therefore, is not the truth of the beliefs we *do* have. We want rather that we *would* believe <*p*> only if <*p*> were true. And from this it does not follow that we want to believe <*p*>, nor does it follow that we want <*p*>. Neither of these follows even on the assumption that we do believe <*p*> and that it is true that *p*. What we desire is only that our beliefs be *safe*; for any given proposition, other things equal we would generally desire this: that we would believe it only if it were true. Desire *neither* for the antecedent *nor* for the consequent is logically entailed by our desire for the conditional. Our general antecedent desire is only for the *safety* of our beliefs, whatever they may be.

However, this general desire for safety is not one that can directly guide how we form beliefs. Compare our desires for delicious meals and for safe walks. These desires can guide us if combined with knowledge of what makes for a delicious meal, and what for a safe walk. But how could we be guided on whether to believe <*p*> by considering whether believing <*p*> would be true? Once we believe that believing <*p*> *would* be true, then (minor exceptions aside) we *already* believe <*p*> and no longer need guidance.

If our desire for safe beliefs (ones that would be true) can provide guidance, it still remains to be seen how, and we shall return to this later.

Perhaps we should relativize to questions already in place. Somehow people get interested in certain questions. They want answers to these questions, *correct* answers. *This* sort of true belief we do value, an answer to one of our questions. Whether we *should* be interested in these questions is another matter that we need not enter in order to understand our desire for truth.

A way to be interested in the truth as such, then, is to be motivated by interest in a question, ranging from questions of gossip, to practical questions of the law court or the legislature, to issues of scientific research. Questions both trivial and momentous can interest someone, whose desire for the truth is then a desire for an answer, a correct answer. Someone who wants the answer to the question *whether p* wants this:

$$\langle \text{If } p, \text{ then } B(p); \text{ and if not-}p, \text{ then } B(\text{not-}p)\rangle[1]$$

That is to say, one desires that if p, then one believes that p, and if not-p, then one believes that not-p.[2]

Again, to be motivated by the truth on a question is to be motivated to believe the *correct* answer; let the chips fall where they may.

C

Does knowledge require responsible believing, believing that is well and rationally guided, and does this require motivation by a desire for the truth as such?

What is it to want the truth *as such*? Wanting the truth as such is not the same as wanting it for its own sake. When we want to know ahead of time whether N is the winning lottery number, we may desire the truth as such even if we do not want it for its own sake, not to the slightest degree. We may want the truth *as such* to some extent even if only as a means to buying the right ticket and getting rich. Compare wanting a key that has a certain brand name etched on it, not *as* one with such an etching, but only as one that opens a certain door. One wants a key with that brand name etched on it, but not *as such*. What one wants as such is rather a key that opens the target door, even if one wants to open that door not for its own sake, but only for gaining access to that room.

Let $\langle c \rangle$ be the proposition $\langle$my deceased parents cared for me at least slightly$\rangle$. Surely this can be known based on good enough evidence. In order to be known, must it be believed out of desire for the truth as such?

Consider this:

Care $\langle \text{If } c, \text{ then } B(c); \text{ and if not-}c \text{ then } B(\text{not-}c)\rangle$

A desire for the truth as such on the question whether c would seem to require adopting *Care* as a desideratum, which in turn involves desiring this: that if not-c then $B(\text{not-}c)$. You might find the belief that not-c so painful, however, that you have no desire whatsoever to have it, not even if it is a fact that not-c. Indeed you would

tend to discount, deny, or ignore any evidence that not-*c*, in a way that makes you less than *epistemically* virtuous on this question. What if you have evidence aplenty for <*c*> and none at all for <not-*c*>? Suppose your epistemically vicious disposition to discount, deny, or ignore contrary evidence is hence never engaged, nor would easily be engaged, given your situation. With all the weighty evidence favoring <*c*> firmly and extensively in place, you are able easily to believe that *c*. *Do you know that c?*

It might be argued that in such circumstances you could not know that *c*, since your belief would not be sensitive to the truth on the question whether *c*, in the following sense:

> One's belief [*B*(*p*)] is sensitive (to the truth) if and only if <*p*> were false one would *not* believe <*p*>.

However, the sensitivity requirement is unacceptable, for many reasons.[3] What we can more plausibly require is not sensitivity but *safety*, the contrapositive of sensitivity:

> One's belief [*B*(*p*)] is safe iff one would believe <*p*> only if it were true.

This more plausible requirement for knowledge will in fact be satisfied in our example, so long as not easily would you believe that your parents loved you without your belief being true. Given enough evidence, as in our example, the condition of safety will be satisfied: In the circumstances, not easily would you believe as you do without being right. Compatibly with this, you might still have believed <*c*> rather than <not-*c*> had it been so that not-*c*.

Compelling evidence might surely enable you to know <*c*> despite being so averse to <*B*(not-*c*)> that you have no desire for *Care*. You might thus know the answer to a question despite having no interest in the truth as such on that question, at least not through desiring the likes of *Care*. So your inquiry and your belief-formation on that question is not really *disinterested*. It is influenced by nonintellectual personal preferences; it is so influenced in that your belief formation would not follow the evidence wherever it might point. On the question whether *c*, certain answers are kept out of bounds by nonintellectual preferences. Even so, if such preferences do *not* come into play, nor would easily do so given your actual situation, if in fact your belief does accord with and derive from the compelling evidence, then you can know in so believing.

If to desire the truth on the question whether *p* is to desire the twofold *D*, therefore, it is implausible that one should be able to know the answer only when one's answer is motivated by such desire. Someone lacking that twofold desire can still know the answer to his question, so long as he accepts it on good enough evidence.[4]

D

We have found no way to understand a *desire for the truth as such* on a given question, or at least not any that would enable requiring guidance by such desire as necessary for knowing the answer. One further account remains unconsidered: Why not understand the desire for the truth as a desire for "matching (conjunctive) pairs" *under that aspect specifically*: "under the aspect of involving a world/belief match of the sort $[X\&B(X)]$." On this conception, one's attitude to a matching pair would embody desire for the truth as such if one desired that pair "under the aspect of its being a matching pair." One might then host three separable "desires" for a given matching pair: under the aspect of involving $<p>$ as a component, under the aspect of involving $<B(p)>$ as a component, and under the aspect of involving a match, of the sort $[X\&B(X)]$, between its two components.

However, this brings us back once more to the waiting room. Suppose we do have a desire for pairs of the sort $[X\&B(X)]$ as such; this is then an aspect of potential pairs that makes them indiscriminately attractive under that aspect, other things equal. If we became aware that, at sufficiently low cost, we could bring one about, therefore, we would want to do so. Suppose our situation in the waiting room to be set up properly, however, so that nothing preponderantly counts against our reaching for the phone book, all other things considered. Would we then want to reach? Even in situations so set up, I for one would not be tempted. Which number is at the top of p. 245? I could not care less.

It might be replied that we do have a desire for matching pairs *of a certain sort* under the aspect of their being matching pairs *of that sort*. All right, but *what* sort? The most plausible sort would be "on questions of interest to us." But this threatens vicious circularity once we consider what it is for a question to be of interest to us, and once we consider the *sort* of interest to us that will be of use to the approach under consideration.[5]

Our twofold desire is supposed to explain what is involved in desiring the truth as such on a question of interest. But it is a *selective* desire, tied to our selective interest in certain questions. Relative to desire for the truth so understood, moreover, one could also define a concept of *disinterested* desire for truth, or desire for the truth not only as such but also to some extent for its own sake. And these definitions would enable us to see how implausible it is to suppose that knowledge requires motivation by any such desire. Knowing the answer to a question does not require that one be motivated by a disinterested desire for the truth on that question, nor even by a desire for the truth as such, whether interested or disinterested.

So we have tried an approach in terms of specific questions and correlated desires for the truth as such. But this approach provides no illuminating account of how the concept of truth does or can have an important role in our intellectual lives or in epistemology.

Perhaps we need a more indirect approach. Perhaps our ideal of truth is one that bears directly *not* on true beliefs but on truth-conducive practices. Perhaps we aim more directly to adopt truth-conducive practices, which can then help us attain truth

and avoid error. What follows will consider this approach. A vicious regress threatens it, I argue, one that can be stopped in either of two ways. One way would appeal to master choices freely and autonomously chosen. The danger here is that this would not give us the sort of truth-connection required for knowledge, nor even for epistemic justification.

The only other approach would seem to involve appeal to virtues constitutive of the thinker's intellectual character or nature: in other words, to deep practices of perception, memory, and reasoning that are both (i) reliably truth-conducive, and (ii) a settled part of the thinker's character or nature.

E

Perhaps we must look beyond the particular belief, and even beyond the particular question whether p, for a motivation toward the truth that could play a role in determining whether a belief amounts to knowledge. Perhaps we should consider the methods, or rules, or policies, or virtues that lead to that belief. Let us adopt the term "practice" as a generic term covering all such items, with no requirement that a "practice" be either social or conscious. How then might truth figure in an account in terms of practices? Here is one way: The subject adheres to a practice to believe propositions that are F, and does so because of a belief that F propositions are true or likely to be true. Adherence to that practice would presumably derive from a further practice of believing propositions that are true. When you believe that p, you are accordingly moved by your desire for the truth so long as your belief derives from such a hierarchy of practices.

Take a Cartesian example where a practice of accepting clear and distinct (C&D) propositions is based on a more general practice to believe what is true, along with a belief that C&D) propositions are true. This may then be combined with the belief *that <3 + 2 = 5> is C&D* to yield the conclusion that this proposition in particular, <3 + 2 = 5>, is true, and to yield in turn belief in it. On this model, behind every fully justified belief lies a practical syllogism whose main governing principle is the practice of aiming for truth.

This account leads to the *reductio* of the waiting room, but that is not its main problem. A second problem is how to conceive of a belief-guiding practice constitutive of our pursuit of truth. Take, for example, the following:

<If <p> is true, then I am to believe <p>>

As a directly guiding principle this is useless. For it can gain purchase on one's conduct only through a belief that <p> is true. But if one believes that <p> is true, then one believes <p> already.

Again, the present account explains how truth bears on our intellectual conduct, by appeal to its supposed place in the major premise of the practical syllogisms that help us guide our intellectual lives, a major premise that hence must lie behind our every justified belief. But this leads also to a regress. How so? Note first that the ex-

planation appeals to a hierarchy of practices held in place by beliefs of the form "Propositions of sort F are true or likely to be true." Only a belief derived from such a hierarchy can be properly motivated and amount to knowledge, or so the account would have it. However, no belief is knowledge if held in place by arbitrary or otherwise defective beliefs. Constitutive beliefs of such a hierarchy, of the form "Propositions of sort F are true or likely to be true," must themselves be knowledge. But these beliefs must then derive from practical hierarchies of their own. And these hierarchies must in turn depend on their own constitutive beliefs, which must then derive from further practical hierarchies. And so on.

The regress threatens to turn vicious since such a hierarchy cannot rise to infinity at the present moment, the moment of truth, when we hold the true belief under evaluation. Suppose we get to the top of a practical hierarchy which at this very moment yields or helps yield that belief. By a topmost practice I mean one *not* itself derived from a further practice then operative. So we assume that some topmost practices will be operative at that moment. Each such practice would seem to be of one of three sorts: either (a) chosen so "freely" as to be arbitrary; (b) constitutive of one's unacquired, deepest nature; or (c) constitutive of one's acquired, second nature, or otherwise acquired. The first option will disqualify any belief based essentially on that practice. The second option takes us back to natural practices part of one's first nature. The third option will raise the question of just *how* that practice *was* acquired and how sustained from its inception to the present. And this will lead to the same set of options, now across time rather than at a time.

It might be argued that a cluster of topmost practices could hold together through coherent mutual support, all at once. So none of them would be *arbitrary*, or absolutely autonomous. But it is absurd that *each* could be wholly legitimated through derivation from the others. The whole cluster would seem arbitrary absent some other consideration that lifts it above alternative clusters. For simplicity, I will therefore treat such clusters as single topmost practices.

Any component of our second nature would be adopted in one of two ways: either it is adopted through some "reasoning" (or "proper exercise of the mind," according to Webster) or it is acquired somehow non-rationally (where the mind is relevantly passive or *im*properly exercised). Either way, we must now evaluate its acquisition (and sustainment). If it is through reasoning that it is acquired, we must evaluate this reasoning. And this leads us to premises (whether invoked at that very moment, or involved in earlier sources), and perhaps to further reasoning in support of *these* premises, and perhaps to further reasoning yet in support of these premises, and so on. I see no way to avoid relying at one or more stages on some component or components of the subject's first nature. And this first nature had better be in proper touch with the truth if what ultimately depends upon it is to be epistemically in good, admirable order.

How far back should one go in assessing present knowledge? Back to habits inculcated through good schooling? Back to one's earliest constitution that eventually joined one's nurture to yield one's adult makeup? Back to the transfer of genes from one's parents to oneself? Back to the evolutionary forces that formed our species? It is not easy to deny any of these a place in the explanation of that truth-connection

that makes for intellectual virtue. It is through such factors, presumably, that one is properly connected to the truth by one's constitution upon reaching years of discretion, or upon reaching the stage of agency and cognition, when one is now a proper doer and knower. At this stage one will have a first nature not owed to one's present or past ratiocination, whose virtue is nevertheless evaluable in epistemic respects, involving how well it suits one for attaining truth worth attaining. By hypothesis this evaluation can no longer depend on prior practices: on prior policies, or methods, or faculties, or virtues.

Even early sources of one's intellectual makeup may be relevant, but we need not always go so far back. Take the Swampman, a being created miraculously by lightning so that he walks out of the Florida Everglades as a fully formed, fully functional contemporary resident of the United States, indistinguishable from ordinary Americans in respects of language, clothing, demeanor, habitual behavior, etc. Is the Swampman properly related to the truth? Much is importantly accidental behind Swampman's current beliefs. Does this sort of accident block his beliefs from being knowledge?

What are we evaluating when we assess beliefs and believers? Is it among other things the qualities that lie behind beliefs? And what is the respect that matters? What do we care about in such belief-yielding qualities? If the evaluation of a quality is epistemic then presumably it will concern how well it suits believers for grasping the truth in certain salient field/circumstance conditions. If so, then we will not be overly concerned with how such qualities got there, so long as they are now a stable part of the constitution of the agent/subject whose belief is under evaluation. But must such a quality take the form of a practice that can count as a motivation voluntarily held by the agent/subject? It is this that seems problematic. If the deepest such practices would not necessarily be held *properly* just in virtue of being deepest, then even for the deepest practices there will be the question of *how* they are acquired and sustained so as to constitute knowledge and so as to yield other knowledge. And *this* proper acquisition and sustainment would not be explicable in terms of practices deeper yet. Nor does such a deepest practice seem properly sensitive to the truth if it is just *chosen* with "*absolute freedom.*" If such a choice is so free as to be arbitrary, how can we understand its connection with the truth? How can we make plausible the notion that the agent chose that practice *because* it is true, and because the agent is so in touch with the truth that he *would* have chosen correctly? If the choice is so "free" as to be arbitrary, how can it be that the agent *would* be likely to select it? This seems mysterious and unsatisfying.

F

What then does constitute our advocacy of the ideal of truth? When is our belief formation virtuously guided by that ideal? If the approach via hierarchies of chosen practices is indeed vicious, and the piecemeal, question-by-question approach problematic, what better approach is there?

Sooner or later we shall need to recognize that our virtuous epistemic conduct

must derive at some deep level from our virtuous nature, a nature not itself due entirely to one's free and autonomous choice. Any choice due to the agent must derive from something in the agent's nature, lest it be unacceptably arbitrary or fortuitous. But that in the agent from which it derives cannot be prior choices unto infinity, even if choices may be affected by temporally prior choices that helped set the agent's character. Requiring a logically prior choice without exception would lead to the vicious regress or to the unacceptably arbitrary. Virtuous conduct must derive from something in the agent's constitution not itself a logically prior choice. If the conduct is really admirable, from an intellectual standpoint, then the constitution manifest in the evaluated conduct must itself be admirable because it helps the subject get into proper relation to the truth. The character from which the intellectually admirable conduct then flows must so constitute the agent that the conduct flowing from it is appropriately enough attributable to that agent as her own, admirably so. At this level, again, that in one's character to which the admirable performance is attributed cannot be some logically prior truth-conducive practice, or policy or motive or virtue.

We have allowed that Swampman's accidental creation as a fully formed subject, complete with underived practices, does *not* disqualify his later beliefs from being properly formed and candidates for knowledge. Yet his original practices are as accidental and arbitrary as can be. How then can we object to a practice derived from absolutely autonomous choice? No such choice is appropriately responsive to the truth. If I just will arbitrarily to accept propositions that are clear and distinct, and if this then gets installed as an underived component of my intellectual character, and if this character stays in place *simply* in virtue of having been thus established and similarly sustained, and not through any later support from other sources, then even if I do accept clear and distinct propositions because I see them to be clear and distinct, my acceptance of them will not manifest appropriate responsiveness to the truth. Note well: The acceptance here is *not* that of a rational being who *would* accept the clear and distinct through self-supporting sensitivity to its clarity and distinctness. On the contrary, by hypothesis one accepts the clear and distinct *only* because of one's arbitrary preference for the clear and distinct. This cannot be the way of intellectual virtue, and cannot yield knowledge.

It does matter, again, just how we conceive of the case. If the arbitrary volition manages to set in place a fundamental practice of accepting the clear and distinct, and if this then stays in place as a self-sustaining and firm part of one's intellectual character, one *not* in need of continuing supportive choices, then our subject *will* be harder to distinguish from the Swampman. To the extent that the underived practice chosen arbitrarily sustains itself firmly and stably enough, and independently of further arbitrary choice, to that extent will it bring our subject's position closer to that of Swampman. To that extent, moreover, will it grow more plausible to grant him intellectual virtue and knowledge.

The choice we face is hence this. We require the subject to be properly constituted so as to "exercise her mind appropriately": that is, normally to accept the deliverances of her senses at face value, to accept what is clear and distinct, to extrapolate inductively in appropriate ways, etc. And we would like all this to be largely a

matter of first or second nature, not requiring case-by-case scrutiny of the sort favored by skeptics. But in what ways can the subject properly acquire her nature? If accidental acquisition is allowed, as with the Swampman, why is acquisition by arbitrary volition disallowed? In my opinion arbitrary volition is disallowed *if* viewed as an ongoing juggle that keeps our practices aloft; but *not*, or not so clearly, if viewed only as an original source of a stable character that then operates on its own, with no further need for juggling by the unfettered will. Why might this be plausible? We are interested in evaluating actions and beliefs for the light this throws on the character of agents and cognizers. We are interested in keeping track of our own and each other's aptitudes and weaknesses, abilities and disabilities, virtues and vices. How surprising can this be in a species as deeply social as ours? What we thus care about, then, is stable, dependable constitution, not arbitrary volitions that can too easily change direction unforeseeably.[6]

Our relation to the environment with whose truth our nature puts us reliably in touch cannot itself be wholly accidental. It can be accidental in some respects. It is accidental that one is alive at all, as accidental as the chance meeting that led to one's parents' marriage. It is accidental that the human species is extant, as accidental as the outcome of the Cuban missile crisis, or as accidental as the physico-chemical combinations within the required range that enabled our evolution. And so on. But these accidents are all compatible with it being no accident, not relevantly, that one is now in an environment, on the surface of our planet, wherein one's perceptual organs are well adjusted to tell it like it is.

My scope in these recent comments has been modest. Beliefs are assessable along several different dimensions. They can be true or false, they can be safe (such that $B(p) \rightarrow p$) or unsafe, and they can be knowledge or not, to take three important dimensions. Here I have focused rather on belief that is "justified" or reasonable, epistemically so. When we categorize a belief thus positively we speak directly of the belief but also, indirectly, of the believer, whose intellectual reliability is also under evaluation. It has been my contention that *one* important requirement on justified or reasonable belief is an appropriate truth-connection, which cannot be secured through essential dependence on absolutely autonomous, free choice. Many sorts of dependence on accident *are* apparently allowed, as in the Swampman case. What one cares about in oneself and in one's epistemic fellows is a relevantly stable, dependable character, however, one whose aetiology is as may be and allows much room for the accidental. What is non-negotiably required to be sufficiently free of accident is the stable, continuing intellectual character of the subject. It is this that makes the subject eligible for epistemic credit and a potential knower.[7]

G

Objections and Replies

Objection 1   It is absurd to grant status as an intellectual virtue to a practice simply because it happens to get installed as a part of one's first nature and happens to be re-

liable. This is precisely what has made reliabilism objectionable. A brain lesion that reliably leads to your beliefs that you have a brain lesion cannot by itself yield intellectually virtuous, justified belief. A palmreader's practice of reading a certain pattern as a sign of serious illness cannot yield an epistemically justified belief if it is an isolated practice adopted arbitrarily, not even if it turns out to be a reliable practice.

Reply   That all seems obviously true. Equally plausibly, however, Swampman can acquire epistemically justified, intellectually virtuous beliefs and can even come to know things when he forms these beliefs through his accidentally acquired practices. So I find myself pulled in opposite directions. How to resolve this? Note first that my suggestions about first nature are *not* a complete account of what is involved in intellectual virtue. Reflecting on our love of truth led me to suggest that truth has a role in the evaluation of practices constitutive of our first nature. But this does not imply, nor do I believe, that *any* reliable practice part of one's first nature constitutes an intellectual virtue. Again, practices need not be acquired *because* they are truth-conducive, not if Swampman's practices are allowable as virtuous. My relevant claim here is only that avoidance of the vicious regress requires that we postulate first-nature practices whose reliability enables them to yield epistemically justified beliefs. This claim does *not* commit us to the view that *any* such practice will do.

Objection 2   It is surely an epistemically important difference between Swampman and normal humans that his first nature is accidentally acquired whereas ours is not. Unlike Swampman, we adhere to our first-nature practices because these are true. Shouldn't our epistemic assessments of the respective beliefs reflect that fact?

Reply   Perhaps we should say that our beliefs are superior to Swampman's simply in being *more deeply* truth-connected than his. And that leads us back to the first objection. The epistemic standing of a practice is boosted not only by its aetiology but also by its coherence with the subject's other truth-connected practices. This would tend to render that practice more securely fixed and safer. Thus supported, the practice would less easily mislead.[8]

Notes

1. Insertion of a declarative sentence within angle brackets will function as a nominalizing device, so that '*<p>*' is tantamount to 'that-*p*'; occasionally capitalization may also function thus.

2. We have focused on the case of yes/no questions of the form whether *p*. But our pattern of explanation fits other questions as well. For example, when we ask who is *F*, our desire for an answer has the following desideratum:

D′   <For all *x*, if *x* is *F*, B(*x* is *F*)>

(Here I assume that if one knows of at least one *F*, that it is *F*, this may answer only partially the question as to who is *F*; a complete answer will need to specify for each *F* that it is *F*. Moreover, to allow for the possibility that nothing is *F*, the more strictly correct desideratum is not D′ but D″: <If something is *F*, then for all *F*, if *x* is *F*, B(*x* is *F*); and if nothing is *F*, then B(nothing is *F*)>. And we should add the following clause, too: that if *x* is not *F*,

then not-$B(x$ is $F)$; otherwise, we would leave it open that we have an incorrect answer to our question. Our desiderata–D and D'— both manifest one's wanting beliefs in accordance with the facts in a certain range, where the range might be just {$<p>$, $<$not-$p>$}, or a set of facts of the form [x is F]. Similar reasoning would seem to apply to questions of other forms, such as which-questions, when-questions, why-questions, etc. In what follows we shall focus mainly on yes/no whether-questions, but our treatment should be applicable *mutatis mutandis* to questions of other forms.

3. There is an extensive literature on this, including *The Possibility of Knowledge: Nozick and his Critics*, Steven Luper-Foy [now Luper], ed. (Rowman and Allanheld, 1987).

4. Is there some other way to understand how it is that one might desire and pursue the truth as such? Perhaps there is a response to our example. Perhaps we do not really need the full twofold desire. Perhaps I do believe that my deceased parents cared for me, and believe it out of desire for the truth as such, despite wanting to avoid belief that they did not even if in fact they did not.

The suggestion is now this. If one believes $<p>$ then one does believe out of desire for the truth as such, so long as one desires $<$If p then $B(p)>$, regardless of whether one also desires $<$If not-p, then $B($not-$p)>$. We next explore this, both as a prop for a responsibilist epistemology requiring desire for the truth as such, and also for the light it might throw in any case on the way or ways in which it is possible to desire the truth as such.

To desire $<$If p then $B(p)>$ is to desire $<$Not-[p&not-$B(p)$]>. (I am of course interpreting the conditional here as material.) To favor *that if* p *then one believe it* is to oppose *that it be so that* p *without one's believing it*. To desire *that if* p *then one believe it* is to desire *that one not miss the fact that* p *if it is a fact*, i.e., to desire *that it not be the case that it is a fact that* p *while one misses this fact*.

One may well prefer that no present headache that one suffers should escape one's notice, for at least this reason: because one prefers that none such occur at all, whether noticed or unnoticed. Trivially, then, one will prefer that one not have a headache while one misses this fact. But such desire out of aversion to headaches is not a desire for the truth as such.

Similarly, I may be quite averse to overlooking my parents' indifference, for at least this reason: because I prefer that it not occur at all, whether noticed or overlooked. Trivially, then, I will prefer that it not be the case that my parents loved me not at all while I miss this fact. But such desire out of aversion to being unloved is not a desire for the truth as such.

How else then might we understand a desire for the truth as such without falling back to the problematic twofold desideratum < if p then $B(p)$ and if not-p then $B($not-$p)>$? Try this:

> On the question whether p, S has some desire for the truth as such if S prefers $<p$&$B(p)>$ to each of $<p$&not-$B(p)>$ and $<B(p)$&not-$p>$.

How plausible is it that preferring $<p$&$B(p)>$ to each of its mismatching alternatives constitutes a desire for the truth as such? What if, as in the case of our example $<c>$ (that my parents loved me some), one has a positive desire for each of $<c>$ and $<B(c)>$ separately? In that case one will naturally desire the matching pair $<c$&$B(c)>$ more than either mismatch, but such desire is not necessarily desire for the truth as such. One does prefer the truth as such in preferring that matching pair only if how much one prefers it is not explained wholly through one's desires for its conjuncts separately. One's degree of preference must exceed what derives from desire for the conjuncts individually.

What would constitute such surplus value? We need some aversion to $<p$ & not-$B(p)>$ that does not derive trivially from the mere aversion to $<p>$ or desire for $<B(p)>$, nor from these in combination. But such aversion to $<p$ & not-$B(p)>$ will not derive from a desire for the truth as such unless it is a general aversion that would be present whatever the content of $<p>$. Compare the desire for a certain key as one that opens a certain door. If this is sim-

ply and solely a desire for a key as such a key, it will entail a general desire for any single such key as such indiscriminately, other relevant things being equal. But this leads back to the discredited account of the love of truth as the desire for true beliefs in general. If we really had such a love of the truth, then we would reach for the telephone book when in the waiting room. There again our aversion to situations of the form $[X\&\text{not-}B(X)]$ would move us to act, given enough time, no prospect of other useful occupation, etc.

Moreover, one has no preference for $<h\&B(h)>$ over both its mismatching alternatives where $<h>$ is the proposition that one has a headache. On the present account one therefore has no relevant desire for the truth as such on the question whether one has a headache. Yet, in spite of this, one's actual headache will surely make its presence known.

5. We have been exploring possibilities for an explanation of what is involved in desire for the truth as such. But so far we have understood the two conditionals in desideratum D as material conditionals.

D $<$If p, then $B(p)$; and if not-p, then $B(\text{not-}p)>$

What if we take these conditionals to be subjunctive rather than material? We can have a single desire for $<$If p, then $B(p)>$ even absent a desire for $<$If not p, then $B(\text{not-}p)>$. A problem with the material understanding of these conditionals is, recall, this: On that understanding the desire for $<$If h, then $B(h)>$ follows trivially from our desire for $<\text{not-}h>$, as in the case of the headache. And a desire for $<$If not-c, then $B(\text{not-}c)>$ follows trivially from our desire for $<c>$, as in the case of our parents' love. But the subjunctive understanding of the conditional has no such problems.

On the subjunctive understanding, then, I might easily desire $<$If c, then $B(c)>$ and yet have no desire for $<$If not-c, then $B(\text{not-}c)>$ despite my overpowering desire for each of $<c>$ and $<B(c)>$. How might we now explain, on the subjunctive understanding, my desire for the truth as such in believing $<c>$? We might now resurrect the "single-desire explanation," saying that my desire for the truth as such in so believing can derive from my desire for $<$If c, then $B(c)>$ singly, and does not require a conjoint desire for $<$If not-c, then $B(\text{not-}c)>$. How defensible is this?

Suppose while desiring $<$If c, then $B(c)>$ I desire with equal intensity $<$If not-c, then $B(c)>$. It follows that I do desire $<$If c, then $B(c)>$. However, my belief of $<c>$ could hardly, in such circumstances, betoken any desire for the truth as such on the question whether c. What I want at all costs, in complete disregard of the truth on that question, is to believe $<c>$. I want to believe $<c>$, with equal intensity regardless of whether it is true or false. So it is not so that the single desire for $<$If c, then $B(c)>$ will sufficiently explain how I might be motivated by a desire for the truth as such in believing $<c>$. We must at a minimum require that this desire not be joined to an equally intense desire for $<$If not-c, then $B(c)>$. Only the combination of the single positive desire for $<$If c, then $B(c)>$ with the absence of desire for $<$If not-c, then $B(c)>$ might therefore give us a basis for attributing to the subject a desire for the truth as such in believing $<c>$.

However, this account seems just as useless as is the full requirement of D itself, if one wants to defend the following view: the view that one can know in believing $<p>$ only if one's so believing is somehow motivated by a desire for the truth as such. Just as someone who does not desire the full desideratum D might still know in believing $<p>$ so long as her evidence for $<p>$ is powerful enough, so someone who desires both $<$If p, then $B(p)>$ and $<$If not-p, then $B(p)>$, showing thereby a certain disregard for the truth as such, might still know in believing $<p>$, so long as she believes on the basis of powerful enough positive evidence and a total lack of negative evidence.

We must be careful with the temporal specifications: One's desire for the truth as such on a question whether p should not necessarily be viewed as just a desire for the truth, at the very moment of desire, of the conjunction of conditionals: $<$If it were so that p, I would be-

lieve that p; and if it were so that not-p, I would believe that not-p>. For example, I can now be interested in the question whether p without desiring that I already be such that the conjunction of conditionals be true of me right now, at the moment of desire for the truth as such on that question. I may not desire that at all, since it would preclude the pleasure of inquiry. What one desires is rather that the conditionals become true at some point, perhaps aided essentially by appropriate inquiry designed precisely to change one through the acquisition of some property F such that, once that property is acquired, the two conditionals are then true of one. But actually even this is not quite right: What one wants may be only that one acquire some property F such that <If p and one had F, then one would believe that p; and if not-p and one had F, then one would believe not-p>. Inquiry on a question whether p may then be understood as an attempt to acquire such a property. Successful inquiry would lead to the fixation of belief through such a property thus acquired (but of course not all processes of acquisition of such a property would properly count as "inquiry").

6. Compare Hume's Treatise (fourth paragraph of Section 1 of Part III of Book III): "If any action be either virtuous or vicious, 'tis only as a sign of some quality or character. It must depend upon durable principles of the mind, which extend over the whole conduct, and enter into the personal character. Actions themselves, not proceeding from any constant principle, have no influence on love or hatred, pride or humility; and consequently are never consider'd in morality."

7. For an alternative take on some of these issues, compare the work of Linda Zagzebski.

5

EPISTEMIC MOTIVATION

Abrol Fairweather

I

A growing number of epistemologists use the language of virtue and vice to eluci-date the concept of knowledge, an approach now known as Virtue Epistemology (VE).[1] VE shifts the focus of epistemic theorizing away from the analysis of familiar concepts like epistemic justification, warrant, and knowledge and toward the con-cept of intellectual virtue. Some VE theorists make this move in order to eliminate troublesome concepts like justification and even knowledge itself, while others use intellectual virtue as a means of illuminating these concepts.[2] These are eliminativist and non-eliminativist variants of VE respectively. In either case, specifying the con-ditions under which a person possesses intellectual virtue, and when doxastic activ-ity is appropriately connected to such a state, is the fundamental task for a VE the-orist. The definitions of justification, warrant, and knowledge come later, if at all.

I think that too little attention has been directed to the nature of virtue itself. It is true that as epistemologists we are interested in just a particular species of virtue, intellectual virtue, but the account of the species is constrained by the account of the genus. That is, what counts as an exercise of intellectual virtue will heavily depend upon the kind of thing that virtue is taken to be. We would thus expect internal disputes between VE theorists over how to define intellectual virtue to lead us to disputes about which general concept of virtue to employ, and for many of these dis-putes to be answered at this level. But there is surprisingly little discussion of the na-ture of virtue itself in VE (a notable exception is Zagzebski [1996]). Virtue episte-mologists appear to take the general nature of virtue either as an issue that has already been settled, or as one which does not require independent philosophical treatment.

I think both claims are mistaken. The Aristotelian conception of virtue as an ex-cellence of character has dominated work in virtue ethics, but that does not imply that it is the best account for epistemologists. There are other plausible accounts available; virtue can be defined as a skill or a mere power (more on these in section

II). When we begin to construct our account of intellectual virtue, we consciously or unconsciously pick one general account of virtue among a field of candidates. We thus cannot take the nature of virtue itself as something that is settled prior to epistemic theorizing, and I think it is best for the issue of choice to be treated as an independent step in the theory.

Moreover, it is philosophically important for virtue epistemologists to look at the different criteria for the possession of virtue that is implied by each of these accounts because this is the material from which we will construct our criteria for the possession of knowledge. An adequate account of knowledge will require that we employ a conception of virtue that allows us to express the full range and complexity of epistemic evaluations. An impoverished account of virtue will leave us with an impoverished account of knowledge.

Here I defend the traditional Aristotelian conception of virtue as best suited to express the full range of epistemic evaluations that matter to us. I will argue that an Aristotelian account captures more epistemic desiderata than either a faculty or skills account. Thus, I defend an account of virtue that requires an agent to possess certain motivational or affective states in order to count as fully praiseworthy. The implication for our epistemic theory is that doxastic activity must be appropriately connected to a certain motivational state in order to be epistemically virtuous. I argue that this desideratum is central to our judgment that an agent's beliefs are epistemically justified or constitute knowledge and that the Aristotelian account is the only one that captures this requirement for virtue. On this basis, I defend the Aristotelian account familiar to us from ethics as the one best suited to meet the evaluative needs of epistemologists.

In section II, I will examine two accounts of virtue that do not require that an agent possess any particular affective state or motivation and provide reasons for taking a closer look at the motivation-based view sketched above. In section III, I will define the nature of epistemic motivation in more detail, and in section IV will argue that whether an agent's beliefs are connected to such a state plays a significant role in knowledge attributions.

II

According to a virtue epistemologist, a person's belief P constitutes knowledge or justified belief if and only if she exhibits intellectual virtue in holding P. This equivalence can characterize either a weak or strong version of VE, depending upon which side of the equivalence is given conceptual priority. A weak VE theorist will accept it, but not because virtue is given any special conceptual or explanatory role in the theory. It just happens to be implied by an independently developed epistemic theory.[3] This is not a very interesting form of VE because virtue and vice play no substantive theoretical role in generating the definition of epistemic concepts or the formulation of epistemic requirements. Weak VE merely uses virtue theory as a novel lexicon for expressing an independent epistemic theory. If virtue is not given any

significant role in an epistemological theory, then it bears only a tenuous relation to *virtue* epistemology, even when the equivalence holds.

The strong VE theorist gives virtue a substantive role in defining epistemic concepts and formulating epistemic requirements because knowledge and justification are carved out by a theory of intellectual virtue. The metaphysical commitments and explanatory models of virtue theory are used to define positive epistemic states.[4] While virtue is given a prominent role here, there may be a limit to how strong our virtue epistemology ought to be. If important epistemic issues are being settled solely by our commitment to a specific form of virtue theory, then our theory might fall more properly under the category of practical rationality, or something outside of epistemology proper.

A virtue epistemologist must therefore not only commit to a model of virtue but to the strength that such a model should have in his epistemological theory. I am interested in a fairly robust form of VE that proceed as follows: (1) Begin with an account of the nature of virtue as a general property. (2) Define the instantiation "intellectual virtue." (3) Define epistemic concepts and requirements under (2). In order to initiate this work, we need to decide what kind of thing a virtue will be in step (1).

Any virtue-based epistemology will evaluate an agent in view of his relation to a certain end. As epistemologists we are interested in evaluating an agent in terms of how she stands to ends that have something to do with truth and falsity. I will call these *alethic ends* because they are individuated by truth modalities, but will leave this notion undefined for the moment. Thus, any virtue epistemologist will define virtue-conferring properties as ones that appropriately relate an agent to an alethic end. But, what do we count as an appropriate relation between an agent and an alethic end? Does the possession of intellectual virtue require an agent to be normatively committed to an alethic end or simply successful in attaining it? Must an agent's standing in such a relation to an alethic end be something for which she bears responsibility? What should be counted among our alethic ends? Different models of virtue lead to different answers to these questions.

Aristotle's view is that virtue is an excellence of character, an enduring state of a person that disposes one to behave in certain ways and also to have certain motivational states directing and controlling this behavior. In *Gorgias* and *Republic* (Bk. 1), Plato defines virtue as an acquired skill—as *techne*. There is yet another way of thinking about virtue that is present in the work of Aquinas (*Summa Theologiae*, I–II, question LV). On this account, virtue is simply the power to bring about a certain end. Faculties like sight and memory are the kinds of things that can possess virtue, but virtue is, properly speaking, the excellence of a faculty rather than the faculty itself.[5]

Each of these models is present in current work in VE. Ernest Sosa (1991) and Alvin Goldman (1978) employ the teleological account, while John Greco (1993) has developed an account that makes virtue an acquired skill. Linda Zagzebski (1996) and James Montmarquet (1993) have taken the traditional Aristotelian approach. Internal questions that arise within an accepted model of virtue are discussed more intensively in this body of work than questions of which model to accept in the first

place. Given the importance of this issue for the rest of a VE theory of knowledge, I think we must examine these meta-questions separately, prior to jumping into steps (2) and (3) above. I turn to these issues next.

Which account of virtue provides the best normative foundation for epistemic evaluation? The teleological account confers virtue upon a thing, process, or capacity so long as it reliably brings about alethic ends. Well-functioning faculties are often seen as those that give us success in reaching these ends. But, whether or not our natural faculties are successful is, for the most part, independent of any skill, effort, or motivation exercised by an agent. Although there are things that we can do to either enhance or diminish our natural capacities, their condition is due largely to natural endowment and the contingencies of our history. Also, to possess reliable faculties does not require that one have a particular motivation or desire.

The strong VE theorist who adopts this general account of virtue will most likely be a reliabilist. Ernest Sosa says, "An intellectual virtue is a quality bound to help maximize one's surplus of truth over error" (1991), and he takes the faculty of sight as a paradigm example of an intellectual virtue. According to Sosa, intellectual virtues are the features of a person that constitute their ability to grasp the truth, grasping the truth being the proper end of the intellect. The epistemically appropriate relation between a person and an alethic end is defined in terms of an objective connection between their belief states and truth. One possesses intellectual virtue if and only if one's cognitive faculties sustain belief-forming processes that are epistemically reliable.

Leaving aside the fact that Sosa defines faculties themselves as virtues, rather than excellences of faculties, this account provides us with slim normative resources because all epistemic values are derived from external success in reaching truth. Sosa recognizes this himself as he adds a deontological conception of justification to his account of knowledge. This is roughly the requirement that we must measure up to our own standards of what is epistemically permissible (1991, 141). I certainly agree with Sosa that some subjective form of epistemic praise is necessary for knowledge, but as VE theorists we must find this within our account of virtue rather than in an independent deontological standard. The faculty account of virtue does not provide these normative resources, leaving us with nothing but reliabilism. Thus, while Sosa is not a reliabilist strictly speaking, this is all we can get out of his faculty account of virtue.

Counterexamples to pure reliabilism are familiar to us by now. Reliable crystal ball gazing and unreliable but rational victims of a demon world both point to the need for a requirement on knowledge that goes beyond external reliability.[6] I find these objections convincing and reject the faculty account of virtue for this reason. But even the steadfast reliabilist, unnerved by such putative counterexamples, may want to enrich his evaluative resources. For example, Alvin Goldman's account of weak justification (1988) requires a richer normative structure than that which the faculty account of virtue provides us. While Goldman does not say that weak justification is a necessary condition for knowledge, an adequate epistemic theory should still have the expressive power to define a normative concept of this kind.

If we conceive of virtue as the possession of a skill, then virtue-conferring prop-

erties must be acquired rather than possessed by nature. Furthermore, their acquisition must be due to some effort made by the agent. The acquisition of a skill requires training and discipline, and each of these requires practical knowledge. On this account, an agent is appropriately related to an alethic end only if his effort and discipline bring it about that he is so related. This makes the agent responsible to some extent for his possessing virtue. Intellectual virtue will be explained in terms of the exercise of intellectual skills such as deductive and inductive reasoning.

Since to possess a skill makes us personally praiseworthy in a sense that is not captured in the faculty account, this approach holds greater promise for illuminating the subjective aspect of epistemic justification. An intellectually virtuous person will now have a degree of understanding concerning the norms that ought to guide her conduct and how such conduct is conducive to bringing about a certain result. Training, discipline, and understanding of a technique are necessary for acquiring a skill, and these in turn imply a set of norms to which we are held accountable. Thus, it appears that we can generate an account of epistemic responsibility from the skills approach but not the faculty approach.

The desideratum of external success in reaching truth (the desideratum favored by the faculty account) is preserved in a skills account because exercising a skill requires external success in reaching the object of the skill. A skilled marksman must succeed in hitting the mark a good portion of the time. The skills account thus has greater expressive power than the faculty account because it preserves the important desiderata of the faculty account and gives evaluative significance to responsibilist desiderata as well. Given the many faces of epistemic praise and blame, this is an advantage.

However, natural endowment can largely diminish both the need for and the utility of an agent's effort. Hitting a baseball requires less effort for people with considerable strength and dexterity, and effort is nearly futile for those with very diminished capacity in these areas.[7] It is true that hitting a baseball is at least as much a matter of proper technique as of natural ability. But two batsmen with the same grasp of a technique and radically different levels of natural ability will most likely not be equally good hitters. This makes the responsibility condition in a skills account fairly weak, and some people will be precluded from virtue by their nature.

One also does not have to be normatively committed to a skill in order to possess it (e.g., as witnessed by the case of a remorseful but accurate bombardier). We might consciously abjure the ends that our skill brings about, yet continue to practice it; perhaps we continue because of the joy that comes in the mastery of the skill itself. This represents a form of practical inconsistency as the agent is akratic in exercising his skill. What bearing this has for evaluating agents epistemically has not been thoroughly investigated at this point,[8] but it strikes me as a defective condition.

The excellence of character account requires an agent to make a deeper contribution to their virtue. Like the teleological account, reliably bringing about an alethic end is necessary for virtue. A benevolent person must in fact contribute to the welfare of others, but we now add as a necessary condition that the conduct flow from an appropriate motivational state. The benevolent act must arise from a desire to improve the lives of others. The possessor of virtue must now be normatively

committed to the ends they reliably attain. This form of agent contribution to virtue is more independent of natural endowment than the kind of contribution implied by the faculty and skills account and thus implies a stronger form of responsibility. The credit generated by having an appropriate motivation appears to be equally attainable by any rational being, and the kind of control we exercise in the instantiation of virtue-conferring properties is stronger in motivational states than on the skills and faculty accounts.

To possess intellectual virtue now requires that an agent's motivational states bear an appropriate relation to an alethic end. What does it mean to be appropriately related to an alethic end? At this point we can say that having an appropriate motivation will be some form of valuing truth, or what I will call an *epistemic motivation* (this concept will be defined in greater detail in the next section). Under strong VE, this commits us to the claim that a belief will count as knowledge only if it is properly connected to an epistemic motivation. Neither the faculty nor skills account has room for expressing this aspect of epistemic evaluation. Thus, I think that the excellence of character account of virtue provides us with the widest range of evaluative tools that we need as epistemologists.

Before proceeding any further I should say something about Linda Zagzebski's recent book, *Virtues of the Mind* (1996). Zagzebski provides a very thorough and well-argued defense of a VE modeled on an Aristotelean virtue ethics. On her account, having a proper motivation is a necessary condition for possessing intellectual virtue.[9] She thus accepts the kind of thesis I defend here. However, Zagzebski has no direct argument for the importance of epistemic motivation in satisfying the conditions of knowledge or justification. This leaves us with the impression that her motivational requirement is a product of the normative framework she uses to define knowledge rather than an essential component of the concept itself. I hope to show that there is a stronger connection between epistemic motivation and knowledge. If I am correct, this would serve to bolster Zagzebski's account.

III

In this section I will develop the concept of an *epistemic motivation.* This will be an affective state of a person; a type of state usually not considered in epistemic theories. A notable exception is Hume, who thought of a belief as just a vivid impression, impressions being in large part constituted by affective states of the person.[10] If we take the Humean view, the very distinction between doxastic states and affective states begins to vanish. Even if we do not go as far as Hume, I think it is undeniable that affective states play a causally significant role in human belief formation. Our cognitive activity is a causally connected tissue of beliefs and desires; desires hence have explanatory significance in cognition.[11]

However, even if we accept this plausible claim about the explanatory significance of affective states, it does not follow that they are epistemically significant. It might be argued that the role that affective states play in cognition is not counted as a necessary condition for knowledge according to our epistemic desiderata. In such

a case, motivational states will be causally significant but not epistemically significant. Let's begin by looking closely at the kind of affective state that is in question.

What is an epistemic motivation? We want to approach an answer by way of the excellence of character account of virtue where virtue is defined roughly as follows: A state of a person is a virtue if and only if it is (i) enduring, (ii) acquired, (iii) a power that allows its possessor to reliably succeed in bringing about a certain end, (iv) a motivation to bring about an end appropriate to the end in (iii). A virtue is a complex state that has a motivation and a disposition to reliably reach a certain end as components.[12]

I will not discuss the requirement that a virtue must be an enduring state in any detail. It is certainly in our favor when a good state is temporally extended, but one might say that a succession of distinct but equally valuable states does us just as well. In his contribution to this volume, Ernest Sosa has a thorough examination of requirement (ii) and argues that virtues need not be acquired. These are important issues that a virtue epistemologist must address, but, as my interest lies elsewhere, I will leave them on the shelf for now.

The requirement of reliable success is important since that alone is what some epistemologists want from the concept of virtue. It appears that a reliabilist has no need for requirement (iv) since virtue as reliability is forthcoming from either the faculty or skills account of virtue, neither of which includes a motivational requirement. Whether or not a reliabilist *needs* something like (iv), it is clear that they need not abjure the epistemic value of motivational states. A reliabilist can make use of an empirically respectable account of motivation and take such states as constitutive features of reliable belief-forming processes without compromising their commitment to naturalism. It may turn out to be an empirical fact about human beings that certain motivational states are components of reliable cognitive processes. Pure reliabilists will not say that having an appropriate motivation is a self-standing source of epistemic value, independent of reliability, but they also should not see an incompatibility between theories that emphasize the epistemic importance of affective states and their own.

What I will examine in detail is the requirement that a person must have an appropriate motivation and that this motivation must play the right role in producing the successful conduct of the virtuous person. This is to add a motivational requirement for the possession of virtue, as well as the possession of knowledge. An epistemic motivation will be a state of a person, but how do we define such a state? What makes such a state motivational in nature? What makes such a state epistemic in nature?

To classify a state as a motivation is to attribute to it the power to influence conduct. Since Hume, we locate that power not in a person's doxastic states, but in a passion or desire. A passion, feeling, emotion, desire, or kindred affective state constitutes the power of a motivational state. Thus, we can say that if some state of a person is an epistemic motivation, it is a state partly constituted by an emotion, feeling, passion, or desire that influences their conduct. For the sake of brevity, I will usually refer to the motivational power as a desire.

What makes a motivation epistemic in nature? A motivation counts as epistemic

not in virtue of something unique to the kind of power it exerts over conduct, but in virtue of something unique to the end it impels us toward. This end will be defined in terms of the intentional content of a certain desire (or set of desires). The type of desire we are interested in as epistemologists is the desire for truth

We can now offer a first approximation to an account of epistemic motivation:

> (EM) A person has an epistemic motivation if and only if he has a desire for truth and this state influences his conduct.

This account will need to be refined. First, the intentional content of the state is given an excessively narrow reading. Clearly enough, to be motivated by the truth is to have an epistemic motivation, but it is not the only epistemic motivation. Earlier I introduced the concept of an alethic end (an end that is individuated by truth modalities). Many ends will count as alethic ends: desiring that one avoid false beliefs; desiring that one only accept true beliefs, or demonstrably true beliefs; desiring to have a comprehensive system of beliefs with a high ratio of true to false beliefs; desiring to have such a system of beliefs now; etc. The category of motivations that count as epistemic should be defined broadly, as is the category of ends that count as alethic. This will include at least being motivated by the coherence of our doxastic system, the reliability of our belief-forming practices, and other justificatory properties.

We could broaden the account further, as Linda Zagzebski does, to include the motivation for wisdom and understanding (1996, 168–176). James Montmarquet (1993) would add open-mindedness and conscientiousness as well. These are all valuable commodities, but one might object that a significant portion of their value comes from the contribution they make to our practical rather than alethic ends, and hence are not sufficiently tied to truth to warrant classifying them as epistemic values.

I do not intend to argue for a specific list of motivations that count as epistemic. Whether we broaden the list as far as Zagzebski and Montmarquet do or just include justification-related values, there is more to epistemic motivation than the desire for truth. Yet, truth still retains pride of place. The reason we value and would be motivated by justificatory properties (or properties taken to be justificatory) is that we think that beliefs that possess them are likely to be true. This connection to truth is what makes us see motivational states individuated by justificatory properties as epistemic in nature. Absent this connection there is nothing epistemic about justification.[13] This suggests that truth should be seen as the primary epistemic motivation from which we can infer derivative epistemic motivations.

While truth is conceptually primary, our immediate intellectual goals—the goals that we explicitly attend to or would cite as the goal of our activity upon reflection—concern truth derivatives more often than truth itself. In the process of inquiry our aim is to follow methods that make our beliefs rational or well supported. The psychological priority of justification and rationality should not be confused with the conceptual priority of truth and should not make us lose sight of the fact that the instrumental connection to truth is what distinguishes justification as an epistemic concept in the first place.

What kind of conduct is influenced by an epistemic motive? Overt behavior can be motivated by a desire for truth. Reading certain books, positioning oneself to get a better visual perspective, declaring a major in philosophy, reading dials in a lab, and re-checking one's work are all overt behaviors that are, or can be, motivated by our desire to increase the number of truths we possess or to increase the likelihood that what we believe to be true is in fact true.[14] But the primary concern of epistemologists is to evaluate beliefs and the cognitive processes that produce belief, not overt behavior. I will follow suit here and focus on beliefs and belief-forming processes, although there is nothing wrong with evaluating overt activity in epistemic terms.

But another question must be raised. (EM) says that a motivational state must influence conduct, and we now understand conduct as belief formation and revision. But how are we to understand "influence"? This is a very tricky question, one that I will not be able to adequately answer here. The trouble we face here is similar to the difficulty faced by philosophers of action in explaining the springs of human action. We could say that a motivation must cause, or be a significant causal feature of, a belief. A simple counterfactual analysis of causation would commit us to something like: S believes P and has motivation M, but if S did not have M then S would not believe P. But, this is too strong because S's believing P may be overdetermined. If there is a counterfactual situation such that S does not have M, but has a non-epistemic motivation R that is sufficient for his believing P, then M cannot be the cause of P even when S has M. But, as Frankfurt has shown, overdetermination is compatible with the claim that one is exercising agency, and the same conclusion should hold for motivation as well.[15] S's belief that P can be motivated by M even if he would believe P in the absence of M.

We could pursue a more nuanced causal account of how epistemic motives influence belief formation. While some causal account seems appropriate here, causality is notoriously recalcitrant to clear analysis and even more so when we are concerned with connections between mental events and the causes of action. I will say that an epistemic motivation influences conduct in the proper sense if it is a salient feature of our explanation of a person's holding a belief.

What makes an epistemic motivation a salient feature of the explanation of a belief? We can begin to appreciate the explanatory role of motivations in belief formation by noting that two people with the same evidence but different motivations can wind up with different beliefs. A person motivated by the goal of holding novel beliefs will respond differently to evidence that makes P likely to be true than would a person motivated to have true beliefs. For the former, the evidence for P would serve as a disincentive to accepting P, since P is the typical thing to believe in the circumstances. For the latter, the evidence for P serves as an incentive to accepting P, since P is likely to be true. Motivations control the course of our transition from doxastic state to doxastic state and hence the set of beliefs we wind up accepting. An epistemic motivation allows us to control the ends that direct doxastic activity, not the doxastic activity itself.

We can now offer the following account of epistemic motivation:

(EM*) A person has an epistemic motivation if and only if they have a desire (or kindred emotive state) for truth or for states whose value is derived from truth, and this desire effectively directs and controls the person's belief formation and revision.

The question I must examine next is why having an epistemic motivation carries epistemic significance. Why should we require that intellectually virtuous persons satisfy (EM*)? What epistemic desiderata does this speak to? Is satisfying (EM*) necessary for being justified or having knowledge?

But before addressing these questions, I will first examine an objection to the very concept of an epistemic motivation. It might be thought that the desire for truth is supposed to function as follows:[16]

(1) *S* desires that the propositions he believes are true.
(2) *S* believes that *P* is a true proposition.
(3) Therefore, desires to believe *P* [as a means of promoting the project stated in (1)].
(4) *S* believes *P*.

We are supposedly lead to (4) as guided by desires (1) and (3). But (2) implies (4) on its own; if *S* believes that *P* is a true proposition, then *S* believes *P* and (1) and (3) are unnecessary in the generation of (4). What is supposed to be an explanation of how we are guided to accept *P* turns out to presuppose that we already accept *P*. This calls into question the very coherence of the desire for truth, as desires are future directed but we are already in possession of what is wanted in the case above.

Step (2) is the one causing trouble, so let us begin there. It is undeniable that once we believe a proposition to be true we believe the proposition itself. But this does not hold for the belief that a proposition is likely to be true, or that there are indications of its truth. A truth indicator can be a proposition, a piece of sensory information, or sets of either that point to the truth of some other proposition or truth bearer. The belief that there are indications of *P*'s truth does not imply that *P* is already believed. In order to produce belief, the truth indicators must be strong enough, but also must be taken by *S* as a significant ground for forming a doxastic attitude of acceptance toward *P*. In an agent that does not value the truth of his beliefs (or not strongly enough), the indications of *P*'s truth will not impel him toward believing *P*. That is why step (1) is important. The desire for truth guides belief formation to the acceptance of *P*, but we have not presupposed that *P* is accepted in the process. Having the general desire that the propositions we believe are true does not imply the belief in any particular proposition's truth, nor does the belief that there are indications of the truth of a particular proposition. If we substitute "*S* believes that there are strong indications of *P*'s truth" for (2), we avoid the objection. Thus, there appears to be nothing incoherent about the desire for truth.

IV

I now return to defending the epistemic significance of motivational states. One thing that can be said in favor of persons motivated by a desire for truth is that, other things being equal, they are more likely to acquire knowledge than persons who lack such motivation. They will be more likely to conduct thorough inquiries, scrutinize evidence carefully, investigate numerous fields of study, consider alternative explanations, etc. In short, the desire for truth gets us to make the kind of cognitive effort that confers justification and knowledge.

This claim is fairly unassailable, as many of our poor epistemic performances are due to a defective level of motivation. True as this may be, it does not follow that having a laudable epistemic motivation is partially constitutive of what makes us justified or that the lack of a laudable epistemic motivation is partially constitutive of what makes us unjustified. The properties that confer justification upon a belief are one thing; the psychological underpinnings that allow our beliefs to instantiate such properties are another thing. Epistemic motivations may be important as enabling conditions of knowledge, but this would not be enough to privilege the excellence of character account of virtue because motivations are here built into the very concept of knowledge.

I will defend the stronger thesis that having an epistemic motivation is part of what makes our beliefs justified and hence part of what makes us knowers. I will consider three cases:

(1) *S* has an improper epistemic motivation guiding his acceptance of *P*.
(2) *S* has no motivation whatsoever guiding his acceptance of *P*.
(3) *S* has good or appropriate epistemic motivations guiding his acceptance of *P*.

In order to defend the importance of epistemic motivation, I will try to show that *S* is epistemically defective in (1) and (2), but not (3). I will also try to show that this kind of defect effects our judgment concerning the justificatory status of *S*'s belief.

Case 1: Belief with Improper Epistemic Motivation

An agent may be said to have bad or improper epistemic motivation if he either (a) shows a disdain for the true and preference for the false or (b) simply shows a lack of concern for truth. The former is a more extreme case of a bad epistemic motive and, I think, borders on incoherence. If *S* desires to have beliefs that are false, he will be motivated by <*P* is false>. But, clearly he must believe (and be motivated by) <"*P* is false" is true> in order for *P* to be a fitting candidate for doxastic assent. *S* now appears to be motivated by the true rather than the false. Or *S* might instead be motivated by the belief that <"*P* is true" is false>. But then *S* must think that <"'*P* is true' is false" is true> in order for *P* to be a fitting candidate for doxastic assent. Again, *S* appears to be motivated by the true rather than the false.

It is not clear that it is even possible to be motivated by the false. In any case, it is also very uncommon to find a person who claims such disdain for the truth. The better case is one where alethic modalities (true or false) are not deemed significant; some non-epistemic motivation directs doxastic activity. Here we have a lack of concern for truth and falsity, not a preference for the false over the true.

Let us consider the case of Conrad, the Doxastic Conformist. Conrad's primary cognitive goal is that a class of his beliefs largely overlap with the beliefs of Mr. Cool. If Mr. Cool believes P, then Conrad will believe P. Conrad forms beliefs in this way not because he thinks Mr. Cool is a reliable guide to the truth, but because Mr. Cool is cool and Conrad wants to be cool. Conrad has become so obsessed with bringing his belief system into conformity with Mr. Cool's that he is no longer sensitive to the alethic properties of his own beliefs or the alethic properties of Mr. Cool's beliefs.

Suppose now that Mr. Cool believes that candidate X will win the election in November and expresses this belief in conversation with Conrad. Conrad immediately adopts the belief that X will win in November. If Conrad does not acquire Cool's evidence for this belief, and has no evidence of his own, then clearly Conrad has not based his belief on evidence and is not justified. To make the case more interesting, let us also add that, through his conformity to Mr. Cool, Conrad acquires Mr. Cool's evidence for the proposition that X will win in November. In conversation, Mr. Cool discloses not only his belief about who will win the election, but also his grounds G for this belief (all the polls show X to be far ahead; the other candidate is embroiled in a horrible sex scandal, is out of money and appears to have suffered a political death already, etc.). Finally, assume that the proposition that X will win the election is objectively probable at the time given G. Conrad now possesses a belief and the justifying evidence for it, but the desire to conform is what directs his belief acquisition. Is Conrad the Conformist justified?

I would argue that he is not. Conrad possesses good evidence G that X will win, but G plays no role *as evidence* in explaining Conrad's acquisition of the belief.[17] The body of evidence he possesses is important to him and effectively guides belief formation, but not because it confers a likelihood of truth on the proposition that X will win. Being supported by good evidence is a purely accidental feature of his belief since Conrad would be just as inclined to believe that X will win without possessing any evidence at all, so long as "X will win in November" is believed by Mr. Cool. Since the properties that make G good evidence are not important for Conrad, it is implausible to claim that Conrad's belief is based on the good evidence he happens to possess.

But suppose that Conrad has independently acquired evidence that X will win, and suppose it is good evidence. His evidence is such that if he were to base his belief on it, he would be justified. Still, this evidence is not mobilized in his belief formation. The evidence he possesses has no bearing on satisfying his desire to conform and is not a salient feature of the explanation of why he holds the belief.

Perhaps, subsequent to his conversation with Cool, Conrad comes to recognize that Cool's evidence G is objectively good evidence for the proposition that X will win. In reviewing their conversation, he correctly identifies the strong evidential relations that obtain among his newly acquired beliefs. While these evidential rela-

tions were not a salient feature in his acquisition of the belief, they may be salient in sustaining his belief. Conrad can now offer reasons for believing that X will win and these reasons can serve as the sustaining ground of his belief. In such a case, one could argue that Conrad is justified in continuing to believe that X will win despite the fact that he acquired the belief with no regard for truth.

I do not think this objection impugns the importance of epistemic motivation. How are we to explain Conrad's attentiveness to evidential relations? He would have a reason to examine his newly acquired beliefs under this guise if he had a concern for their truth, but this commits us to attributing an epistemic motivation to him. It could be that he has no special desire for truth but just happens to recognize that one of the beliefs he acquired is given strong rational support from other beliefs he acquired. But this does not explain why this recognition would sustain belief. Without an epistemic motivation, the recognition that one of his beliefs has positive epistemic status will not give him any reason for sustaining a commitment to it.

If Conrad continued to believe that X will win on the basis of G even if Mr. Cool changed his mind and came to believe that Y will win, then clearly the evidence sustains his belief. I have no objection to saying that Conrad is justified in such a case. But it only seems plausible to say that he is justified because the desire for truth seems to be at least a sustaining element of his belief. If Conrad would come to believe that Y will win as soon as he realizes that Mr. Cool believes that Y will win, while still recognizing that the evidence favors X as the winner, then the evidence he acquired is not sustaining his belief. In such a case, the desire for truth is not effective in Conrad.

I think the case of Conrad the Conformist shows that the mere possession of good evidence is not itself sufficient for that evidence to justify a belief. The evidence must be engaged by our belief forming practices *as evidence*. This means that the properties that make a body of evidence good evidence must be seen as important enough to initiate and direct belief formation. This, in turn, requires that we have a desire for truth and that this desire is effective. Without an appropriate epistemic motivation, evidence cannot play an explanatory role as evidence in our belief formation. I think it is right to conclude that it cannot play a justificatory role either.

Before moving to the next case, one might ask: What if conforming to Mr. Cool is a generally reliable process of belief formation for Conrad? I would argue that that does not add anything new, or of additional value. It will still be the case that evidence is not playing a significant role as evidence in belief formation. What makes such cases epistemically defective has nothing to do with the likelihood of having true beliefs by following such a policy. The defect is that the explanation of Conrad's beliefs does not mobilize evidential considerations, and this will hold whether or not his belief-forming processes turn out to be reliable.

Case 2: Non-Motivated Belief

We have seen that a lack of concern for truth is sufficient to make a believer unjustified, even when he has good evidence for a belief he accepts. This supports my claim that having an appropriate motivational state is a necessary condition for

knowledge. A harder case would be where a believer has no affective states (as opposed to improper affective states) connected to a belief. Consider a being, Robotic Robert, who has impeccable evidence for his beliefs but who has no motivational states guiding his belief formation. Robert's belief formation is driven by evidence and results in a high ratio of true to false beliefs, but no motivation, epistemic or non-epistemic, is involved in the process. In a robot-like fashion, Robert forms beliefs according to the rule "Adopt an attitude of acceptance toward a proposition P if and only if I have evidence that strongly indicates that P is true," and this results in a high ratio of true to false beliefs.[18]

Ex hypothesi, Robert is not driven by appropriate epistemic motivations, because no motivational states guide his belief formation. Yet Robert's performance is, from an epistemic point of view, quite good. One response is that we are left without an explanation of why Robert forms beliefs according to the above rule rather than some other rule and why the evidential properties of propositions matter to him. I will return to this later, but it may be claimed that the point of the example is precisely to show that we do not need to know this in order to reach a judgment on the justificatory status of his beliefs. Robert just forms beliefs according to certain rules, and these rules give salience to evidence in guiding belief formation. Does Robert not appear to be justified despite lacking proper epistemic motivation?

One response is that Robert may not have any grasp of the evidence that drives his belief formation or have any sense that he forms beliefs according to the above rule. If Robert's conforming to the rule is unconscious, or Robert is some type of computer, then he (or it) will not be able to give an account of why he holds the beliefs he does. Being justified in holding a belief should require at least the ability to provide a reason for holding it when queried. Suppose that Robert is able to cite the above rule as the reason for holding a particular belief. Yet if we were to ask, "Why is having indications of the truth of a proposition a good reason to adopt an attitude of acceptance toward it?," Robert would be hard pressed to provide an answer. He could say that, as a matter of fact, he forms beliefs according to such a rule. But he was asked why he ought to, not why he does. Robert has no sense that truth is the right or appropriate aim for belief and it thus seems odd to attribute justification to the beliefs of such a creature.

I am willing to concede that it is not entirely implausible to say that Robert is justified, but I think this really serves to reinforce the importance of epistemic motivation. What makes us willing to see his beliefs as justified is that the rule of belief formation he follows automatically picks out the alethic properties of propositions as salient. But human beings are not like Robert. The properties that guide our activity (doxastic or overt) are the ones privileged by our desires. If the desire for truth is not guiding us, then some other desire will take the helm, say the desire to conform or attain power over others, and this will give salience to the non-alethic properties of propositions. If we can manage to mimic Robert's epistemic performance, this is because the desire for truth is effective in us.

Case 3: Belief with Good Epistemic Motivation

Given the above discussion, the description of this kind of case is fairly evident. We have a believer, call her Barbara, whose belief formation is motivated by a desire for truth. Suppose Barbara has just received a marriage proposal from a young man who produces a large ring and places it on her finger. She knows that a diamond is the customary stone for such a ring, has no reason to believe the young man would flout convention, and is able to distinguish diamonds from other gems and cheap imitations. Barbara comes to believe that she is wearing a diamond ring on her finger. She is also justified in holding this belief.

Our concern here is not the evidential merits of Barbara's belief. Both Conrad and Robert had good evidence for their beliefs, but this was not sufficient for being justified. Barbara also has good evidence, but, unlike Conrad and Robert, she is motivated by the desire for truth. Barbara's desire for truth gives salience to evidence *as evidence* in her belief formation, and this allows her to avoid the kind of epistemic defect that Conrad and Robert suffered from. This difference is also what leads us to judge that Barbara, but not Conrad or Robert, is epistemically justified. I think these examples show that having an appropriate epistemic motivation is a necessary condition for knowledge.

It may be objected that the scope of this claim is too broad. A proper epistemic motivation appropriately connects a belief to a body of evidence, but not all beliefs are based on evidence. Perceptual beliefs appear to be acquired in a direct, non-reflective, passive manner rather than a process of inference from evidence. Thus it appears that epistemic motivations have no important role to play in the justification of perceptual beliefs. If this is the case, then there is a large and important class of beliefs whose epistemic status does not depend on the relation they bear to an epistemic motivation.

I think this line of thought is mistaken — epistemic motivations do have importance in the justification of perceptual beliefs. It is true that our perceptual apparatus can perform accurately even if the desire for truth has no influence over its workings. But, as Ernest Sosa argues (1991, 240–241), perception in rational beings does not produce belief independently of reflective cognitive processes. I think Sosa is correct here and that his argument also shows that epistemic motivations are relevant to the justification of perceptual beliefs (although I do not know if he would accept the latter claim).

According to Sosa, one has *animal knowledge* that P if P is a direct response to sensory impacts with little or no benefit from reflection or understanding (Sosa, 240). One has *reflective knowledge* that P if "one's judgment or belief (that P) manifests not only direct response to the fact known but also understanding of its place in a wider whole that includes one's beliefs and knowledge of it and how it came about." The difference between these two types of knowledge is that reflective knowledge requires that our beliefs have a certain agreement or coherence with other beliefs, whereas animal knowledge can be attained independently of any horizontal connections among beliefs. No additional cognitive process is necessary over and above perception for animal knowledge, but reflective knowledge

requires perception plus a certain understanding of the origin of our perceptual reports.

It is plausible to say that what results from a direct response to sensory impacts is not affected by the presence or absence of the desire for truth. This is a matter of our perceptual hardware and how it reacts to the pressures of its environment. While states of desire will interact with sensory states in a cognitive system, the latter appear to have epistemic standing independent of the former. Thus, if human perceptual beliefs are instances of animal knowledge, then their justification may indeed be independent of epistemic motivation.

Sosa argues that all human knowledge is reflective knowledge, and I am in agreement with him here. Sosa says:

> Note that no human blessed with reason has merely animal knowledge of the sort attainable by beasts. For even when perceptual belief derives as directly as it ever does from sensory stimuli, it is still relevant that one has *not* perceived the signs of contrary testimony. A reason-endowed being automatically monitors his background information and his sensory input for contrary evidence and automatically opts for the most coherent hypothesis even when he responds most directly to sensory stimuli." (240)

I think Sosa is right that perceptual beliefs are produced through two distinct cognitive processes, the mechanisms of perception and a background monitoring system. But the kinds of cues we look for in our environment and in our perceptual apparatus may not be keyed to preserving the truth conduciveness of our sensory reports. We may seek to insulate certain beliefs from falsification and hence ignore cues that suggest their falsity. This would still be a monitoring system, but, from an epistemic point of view, it would be a defective one. Thus, we cannot simply take it for granted that the monitoring system in reflective cognition is keyed to preserving the truth conduciveness of our doxastic system. Our monitoring system must be governed by a desire for truth, or a kindred alethic end, in order for it to function as Sosa describes.

Sosa makes the important point that the credibility of sensory reports can always be overridden by background information, and even when we accept the reports of our senses our monitoring system is a silent partner in producing belief. But our monitoring system only makes a positive contribution to the epistemic status of our beliefs when it is controlled by a desire for truth. I think this shows that epistemic motivations play a significant justificatory role even in the case of perceptual beliefs.

To sum up, I think it is undeniable that Barbara is a more praiseworthy epistemic subject than either Conrad or Robert. I have tried to show that the kind of praise that we give to Barbara but not to Conrad or Robert is closely connected to epistemic justification. The issues discussed above become complicated when we recognize that practical desires will inevitably have some influence on belief formation. It would be overly stringent to require that truth be the sole motivation in our cognitive lives. It seems likely that the desire for truth grows out of the instrumental connection between having truths and satisfying our practical desires;[19] the desire for

truth then takes on a life of its own and attains it a greater or lesser role in directing our activity. Leaving these difficult issues aside, I hope to have shown that having an appropriate epistemic motivation is part of what turns true belief into knowledge. I think this provides us with sufficient reason to adopt the Aristotelian account of virtue for epistemic theorizing.

V

Virtue epistemology is a rich and growing field of philosophical inquiry. Contemporary epistemologists like Ernest Sosa, Linda Zagzebski, John Greco, Alvin Plantinga, Christopher Hookway, and others have begun the important work of cashing out the fertile concept of intellectual virtue. I think it is important to look closely at the different conceptions of virtue before committing to a specific account of intellectual virtue. The classical Aristotelian conception of virtue that dominates work in virtue ethics requires, among other things, that an agent have an appropriate motivational state in order to possess virtue. Of course, this does not automatically imply that the Aristotelian account of virtue is the best account for epistemologists. I have argued that having an epistemic motivation is a necessary condition for knowledge. If I am correct, then it turns out that the model of virtue best suited for moral evaluation is also the one best suited for epistemic evaluation.

There are issues that remain unresolved even if this thesis is accepted. Does a motivational requirement apply to all types of knowledge? We will want a more definitive account of which motivational states count as epistemic. There are important questions remaining about the connection between epistemic and pragmatic motivations. We will want to know exactly what kind of relation must obtain between an epistemic motivation and a belief.

Very little has been written by contemporary epistemologists about motivation as a condition for knowledge, and more needs to be said than is said here. I believe that epistemic motivations provide a psychological link between belief and evidence that is necessary for the possession of knowledge. I hope to have at least provided the impetus for further pursuing the significance of motivational states in epistemology.

Notes

1. Linda Zagzebski (1996), Ernest Sosa (1991), John Greco (1993), Alvin Plantinga (1993), James Montmarquet (1993) all have theories of knowledge that build an aretaic form of normativity into the definition of epistemic concepts. Christopher Hookway and Robert Audi each defend a virtue-based theory of knowledge in their contributions to this volume. There are many different interpretations of virtue among this field of authors, but little direct comparison of the different interpretations. The question of the nature of virtue itself is not treated as an independent epistemic issue.

2. Ernest Sosa does not seek to reduce all epistemic concepts to virtue concepts, as his account of justification is deontological, whereas Zagzebski's stronger version apparently

does. Zagzebski, but not Sosa, uses virtue to eliminate justification from our epistemic lexicon, but she does not intend to eliminate knowledge.

3. Plantinga's Proper Function Theory (1993) may be the best example of weak VE. He does not mind being called a virtue epistemologist, but he has no special allegiance to the title. It turns out that his theory can be nicely captured in a virtue framework. Goldman (1978) might be an example as well. Simon Blackburn appears to accept a strong version of VE in his contribution to this volume, as truth itself is defined in terms of intellectual virtue. Linda Zagzebski clearly defends a strong form in (1996).

4. One significant change is that our epistemic theory no longer takes beliefs as the fundamental unit of analysis. Enduring states of persons, dispositions of character, are now the primary object of epistemic evaluation. For a thorough discussion of this important shift see chapter one of Zagzebski (1996).

5. There is a discussion on this topic between John Greco and Linda Zagzebski in *Philosophy and Phenomenological Research*, March 2000. Greco argues that, according to Aquinas, a virtue is a power, and Zagzebski argues that, according to Aquinas, virtue is, properly speaking, the excellence of the power.

6. See BonJour (1980) for the first example; Cohen (1984) for the second example.

7. I here borrow from an example of John Greco's (1993), although he does not use it to make the same point.

8. Christopher Hookway provides a thorough examination of epistemic *akrasia* in his contribution to this volume.

9. Zagzebski (1996: 270) defines knowledge thus: "Knowledge is a state of cognitive contact with reality arising out of acts of intellectual virtue." An act of an intellectual virtue A is "an act that arises from the motivational component of A." Hence knowledge requires motivation on her view.

10. See Hume, *A Treatise On Human Nature*, Bk I. 3.7

11. For more on how affective states are involved in cognition see chapter 1 of Zagzebski (1996). Hilary Kornblith has an interesting account of how desires function in directing our cognitive lives in (1993). Kornblith argues that much of human cognition is concerned with determining means and this in turn makes truth an instrumental value for any rational creature.

12. This is essentially the account of virtue that Zagzebski uses in (1996). She also discusses Eudaimonistic accounts of virtue in Aristotle that imply a different analysis of intellectual virtue.

13. This is Stewart Cohen's argument in "Truth and Justification" (1984).

14. Theories that evaluate overt conduct and the social environment influencing such conduct are forms of social epistemology. See Goldman (2000) and Corlett (1996).

15. Frankfurt's argument is in "Alternate Possibilities and Moral Responsibility" (1969). Linda Zagzebski examines the significance of "epistemic Frankfurt cases" in "Must Knowers Be Agents?" (this volume).

16. An objection of this kind is suggested by Sosa in "For the Love of Truth?" (this volume), although I am not sure that he intends the specific form I have given it here.

17. Conee and Feldman (1985) make the distinction between having evidence and having evidence as evidence. It is reasonable to say that Conrad possesses evidence, but not as evidence. Feldman pursues the concept of having evidence further in (1988), but does not provide a detailed analysis of having evidence *as evidence*. I think this requires having a desire for truth, but I will not be able to pursue this here.

18. I would like to thank Francis Dauer for suggesting this example.

19. For an extended discussion of such an account see Kornblith (1993).

References

Aquinas, Thomas. 1947–48. *Summa Theologica*, 3 vols. Fathers of the English Dominican Province (trans.). New York: Benziger Brothers.

Aristotle. 1985. *Nicomachean Ethics*. T. H. Irwin (trans.). Indianapolis: Hackett Publishing.

BonJour, Lawrence. 1980. "Externalist Theories of Empirical Knowledge.*"* In *Studies in Epistemology, Midwest Studies in Philosophy*, edited by Peter A. French, Theodore E. Uchling, and Howard K. Wettstein. Vol. 5., pp. 53–74. Notre Dame, Ind.: Notre Dame University Press.

Cohen, Stewart. 1984. "Truth and Justification," *Philosophical Studies* 46:279–295.

Corlett, Angelo. 1996. *Analyzing Social Knowledge*. Lanham: Rowman and Littlefield.

Feldman, Richard (with Earl Conee). 1985. "Evidentialism," *Philosophical Studies* 48: 15–34.

Feldman, Richard. 1988. "Having Evidence." In *Philosophical Analysis*, D. F. Austin (ed.), 83–104. : Kluwer Academic Publishers.

Foley, Richard. 1993. *Working Without A Net: A Study Of Egocentric Rationality*.
Cambridge: Harvard University Press.

Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy*, 66 (Dec.): 23.

Goldman, Alvin. 1978. "Epistemics: The Regulative Theory of Cognition," *Journal of Philosophy*, 75 (Oct.): 509–23.

———. 1988. "Strong and Weak Justification," in Philosophical Perspectives, 2, Epistemology, edited by James Tomberlin. Atascadero, CA: Ridgeview Publishing Co.

———. 2000. *Social Epistemology*. New York: Oxford University Press.

Greco, John. 1993. "The Virtues and Vices of Virtue Epistemology," *Canadian Journal of Philosophy* 23: 413–32.

Hume, David. 1967. *A Treatise Of Human Nature*. Oxford: Clarendon Press.

Kornblith, Hilary. 1993. "Epistemic Normativity," *Synthese* 94: 357–76.

Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, Md.: Rowman and Littlefield.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.

Sosa, Ernest. 1991. *Knowledge in Perspective*. Cambridge: Cambridge University Press.

Zagzebski, Linda. 1996. *Virtues of the Mind*: *An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge.* New York: Cambridge University Press.

6

EPISTEMIC VIRTUE AND
JUSTIFIED BELIEF

Robert Audi

A major presupposition of ethics is that our moral lives can be con-
ducted well or poorly. Our intellectual lives — our questioning and
judging, our reflection and inference, our criticism and responses to criticism — can
also be conducted well or poorly. We value moral virtues as features of character that
govern our overall behavior; but there are also intellectual virtues, such as open-
mindedness and rigor, that may govern much of our cognitive behavior. Virtue
ethics has always been an important theoretical approach in moral philosophy and,
in the past few decades, has received renewed attention; it should be no surprise that
virtue epistemology has also been explored and developed in recent years.[1] Virtue
epistemology is interesting and important in its own right, but there is also value in
comparing it with virtue ethics. Does it have a similar range of resources and prob-
lems? And are there important points still to be learned from the right kind of com-
parison? I think the answer is "yes" in both cases, and my main effort here is to
make some comparisons and bring out some difficulties for virtue epistemology in
at least some of its forms. I'll start with the concepts of (non-intellectual) virtue and
of action from it, proceed to the concepts of epistemic virtue and of belief from it,
and then explore some major connections between epistemic virtue and, on the
other side, justified belief and knowledge. The last section will consider the extent
to which the concept of epistemic virtue provides a basis for understanding justifi-
cation and knowledge independently of a more particularistic, or at any rate non-
trait-based, approach.

I. Action from Virtue as a Model for
Epistemically Virtuous Belief

In the common sense in which virtues are admirable traits that above all guide
action, I take it that broadly speaking a virtue is a feature of character that has a

significant tendency to influence conduct and supplies its possessor both with normative reason indicating what sort of thing should be done in a wide range of contexts and with motivation to do such things for an appropriate kind of reason. An honest person, for instance, not only tends to be truthful, but to do so for a certain kind of reason, say in order to realize a standard of conduct, as opposed to avoiding reproach. The standard can be an ideal, and the reason need not come from a rule or principle; the agent may be guided by an effort to be like a role model. But virtue is not a mere capacity for good deeds. It is (at least) a settled tendency to do them for an appropriate kind of reason. The range of reasons is wide but is unified by the domain of conduct governed by the virtue and what it calls for in that domain. With this notion of what constitutes a virtue goes an account of what it is to act *from* a virtue. Let us consider that and then explore how the notion applies to epistemic virtue in relation to belief.

On the broad conception of acting from virtue I find most plausible — and here I do not claim to be drawing on any particular virtue theory — it is aretaically grounded intentional action: action grounded in virtue either directly, as where the agent acts explicitly in the light of the concept of the virtue in question and with a commitment to the relevant standard(s), or indirectly, as where one acts on the basis of a different kind of consideration that is nonetheless suitably relevant to the virtue in terms of its field and target.[2] The first case is illustrated by a virtuous agent's making a just distribution of fellowships with academic justice as a goal rooted in the agent's character; the second would be illustrated by doing this with the merits of each candidate in mind in relation to the others and a due sensitivity (again rooted in one's character) to the details of each case in relation to the whole, yet without aiming at or being guided by an overall concept of justice.

In either case the action is performed *from* virtue in being explained by beliefs and desires properly connected with the appropriate aretaic elements in character; and it is *virtuous* both because of its connection with those cognitive and motivational elements and because of what kind of action it is (a just distribution). In both cases, one might say, there is an appropriate good that functions as the *telos* of the virtue (roughly, the end — possibly many-faceted — that its exercise properly serves). Avoiding evil can also be an appropriate telos and may be independent of promotion of a positive good (a point easily neglected if, naturally enough, one thinks of virtues as aiming at something good).

One further important point is that in any case of acting from virtue, it is essential that the action be guided in the right way by elements of character, including both motivational and cognitive elements. The mere achievement of an appropriate result is not sufficient. Giving a deserving candidate a fellowship by accident is not acting from any virtue; nor is giving it quite designedly if the deed is either self-serving or produced by some *other* agent's manipulating one's choice. The aretaic notion of a telos, then, must be understood in a sense that is internal at least in requiring an appropriate basis of the action, rather than in a sense that demands only external realization of a good.

II. Epistemic Virtue

To what extent is epistemic virtue analogous to moral and other "practical" virtues? We need some examples. It is not obvious where they should come from. The term "epistemic virtue" is a philosophical coinage. If we are guided by the analogy to ethics, we might fruitfully look to epistemological theories corresponding to major ethical views that offer or imply an account of moral virtue. For utilitarianism, a moral virtue should be understood (roughly) as a trait whose expression conduces appropriately to the proportion of happiness to unhappiness in the relevant population (though this is not how utilitarians have generally represented moral virtue).[3] For Kantian and intuitionist deontological theories, moral virtues are traits that represent a kind of internalization of moral principles, including an appropriate disposition to deal with conflicts of duty that sometimes arise from the applicability of more than one such principle — if practical wisdom is itself a virtue, it may be regarded as in part constituted by this disposition.[4] For an Aristotelian virtue theory, moral virtues are (roughly) traits of character whose expression (or "realization," in a sense) is *constitutive* of an excellent life in the domains in question (these domains are not sharply separable from non-moral ones, particularly given Aristotle's theory of the unity of the virtues). It will be useful for our purposes here to draw some analogies between ethics and epistemology, beginning with rule ethics and proceeding to virtue ethics, in each case considering an aretaic version of the epistemological position in question.

The closest epistemological analogues of utilitarian theories are various forms of reliabilism, with true belief playing a role parallel to that of optimal action and, for an aretaic version of the view, serving as the telos for the relevant traits. The idea is very roughly that just as right action is action that is most likely to produce the greatest good relative to its alternatives, justified belief is produced in a way that makes it superior to its alternatives (including withholding the proposition) in likelihood of truth. Action that actually does optimize goodness is objectively right; belief that is produced in the appropriately reliable way and is objectively true and constitutes knowledge.[5] The direction of causation is different — from mind to world (as it is sometimes put) in the ethical case and from world to mind in the epistemological one — but the role of causation is similarly crucial in both cases.

The closest epistemological analogues of Kantian and intuitionist theories, by contrast, are various kinds of epistemic deontologism, with something like epistemically responsible belief serving, for an aretaic version of the position, as the telos. There are many possibilities here. One is that there are certain rules, such as rules requiring the having of good evidence, and conformity to these rules will determine when a belief is justified. There is controversy regarding how much voluntary control of belief we need for such a theory to be used as a realistic guide in our cognitive life, but that issue may be left aside here.

The analogue of Aristotelian virtue theory is a more "agent-centered" epistemology, with traits of intellectual character, such as conscientiousness and open-mindedness, construed as central for understanding justification and knowledge.

Thus, it is not enough that a belief be grounded in a reliable process in an agent; a justified belief must in some sense reflect character.[6] The *agent* must be justified, by virtue of some feature of intellectual character, in holding the belief.

This essay does not single out any of these kinds of virtue epistemology, but what I say may bear on all of them. In part because I presuppose none of them and indeed explore some difficulties for virtue epistemology in general, my selection of examples of epistemic virtues is not tied to any of them. Fortunately, the notion of a virtue as a trait of persons is not technical; and even if the concept of the epistemic is semi-technical, some of the paradigms of what might be considered intellectual as opposed to moral virtues should also be clear cases of epistemic virtues. In seeking examples, we should bear in mind that good (including "successful") cognitive behavior includes both appropriate formation of non-inferential beliefs and suitably building on these by various kinds of inferences and hypothesizations. With this in mind consider open-mindedness, rigor, and insightfulness. These are not only admirable intellectual traits, but they nicely contrast with the vices of narrow-mindedness, intellectual laxity (roughly "looseness"), and, in the third case, such defects as being superficial, dull, and unimaginative (there may be no genuine antonym of "insightfulness," but "blind" in one use and the uncommon term "imperceptive" come close). All of these terms apply to patterns of non-inferential belief formation, but rigor as opposed to laxity is a trait that governs inferential belief formation at least as much as the non-inferential cases, whereas open-mindedness and insightfulness are perhaps more commonly manifested in forming non-inferential beliefs.

If we take an Aristotelian aretaic perspective on the notion of epistemic virtues, we might naturally try to identify a kind of mean that such virtues should hit. There may, however, be no one dimension on which all of these virtues fall. For one thing, there seem to be intellectual virtues that are not a matter of responding to evidence or grounds for belief, though these *virtues of response*, as we might call them, are the most common. There are also *virtues of pursuit*, such as intellectual curiosity, and *virtues of production*, such as (intellectual) creativity. If there is a mean for the virtues of response, the best candidate for it is roughly a mean between skepticism and credulity, which we might conceive of as the extremes. Too much openness of mind and we get credulity; too little and we get skepticism, where — in my view — this goes beyond being rigorous. I grant, however, that the distinction between virtuous rigor and excessive skepticism is a perennially disputed demarcation. In any case, we need something different for curiosity and creativity, even if skepticism and credulity are relevant to both.

It might seem that our list of representative virtues could be pared down to the virtues of response. I doubt this. Again, the analogy to ethics is instructive. Don't we need to recognize and indeed promote an epistemic virtue that plays a role like that of the moral virtue of beneficence as a tendency to pursue a general kind of good for persons? Neither open-mindedness nor rigor nor even insightfulness requires pursuing truth, as a beneficent person must pursue the good for others. They are in a sense passive, operating on stimuli that in one way or another are brought to the intellect. The counterpart of beneficence is something like (intellectual) curiosity. It is a kind of intellectual motive that leads one to seek to learn things of a certain range.

Without it, moreover, there would be less to respond to and, very likely, less knowledge. Without creativity, there would likely be both less knowledge and certainly less insight.

The ethical analogy also invites us to formulate counterpart deontological principles. Speaking globally, just as we ought to avoid doing evil and seek to do good — and indeed to try to have these two quests interact in a mutually supportive way — we ought to avoid believing (significant) falsehoods and seek to believe significant truths and to try to have these two efforts interact in a mutually supportive way. Truth and falsehood function, positively and negatively, as a telos for cognition, much as good and evil function as a telos for action. There is, however, a disanalogy, since belief is not action and — apart from at most a limited range of cases — not voluntary.[7] How serious is this disanalogy for those who favor a virtue epistemology?

It is not, I think, crippling. For one thing, we do have a good measure of indirect voluntary control over belief formation, for instance by exposing ourselves to evidence and by making a determined effort to resist forming beliefs when we find ourselves lacking certain kinds of grounds.[8] This is part of epistemic virtue: in a well-developed form, at least, it requires that one know what kind of proposition asserted by someone, or otherwise presented for acceptance, is in need of grounds, and what kind of grounds it needs; and it requires that one also be able to resist coming to believe such propositions unless one has grounds. For another thing, it would surely not be desirable to have direct control of belief formation in anything close to an unrestricted way: If belief formation is to guide us in navigating our environment, its formation and alteration must be by and large a natural and reliable response to the world. Even acts of causing oneself to believe something (or to continue to believe it), as where this is necessary to stave off depression, need the guidance of beliefs that are not formed at our pleasure and are resistant to alteration by our felt needs. If people afflicted with cancer have no genuine evidence for thinking they can continue their work better if they come to believe the disease is not serious, then they should not expect producing that belief to help them. If the motivating belief is not evidentially well grounded but either a product of wishful thinking or a result of a previous "decision to believe" not based on adequate evidence, it is not reasonable, on the basis of that belief, to induce in oneself the optimistic belief that seems needed to facilitate one's work.[9]

It might seem obvious at this point that, like moral virtue, epistemic virtue is intrinsically good, in the minimal sense implying that it is good in itself and not merely an instrumental good.[10] But this is not obviously so. In part because so much of belief formation, and indeed so many aspects of belief, such as its strength and its influence on other beliefs and on conduct, is involuntary, the connection of belief formation with a properly guided will is less direct than the counterpart connection in the case of moral virtue. We *do* the things that above all manifest moral virtue; and by deciding against them we can restrict and in a way nullify or — if we form sufficiently bad habits — even ultimately uproot the virtue. But we do not do our beliefs, nor can we effectively decide to be less rigorous or open-minded or insightful (as opposed to deciding to act in the ways that less rigorous, open-minded, and in-

sightful people do). There is, then, a stronger case with epistemic virtues than with moral virtues for saying that although they are good as a means—in this case to getting us around in the world or, in any event, getting us into the good state of cognizing truths—they are not good in their own right. Still, the contrast I am drawing should not be exaggerated. Epistemic virtue is an essential aspect of a well-functioning person and is essential to intellectual flourishing in human beings. To that extent, at least, it surely is an intrinsic good.

This issue raises a question that has so far been in the background. Supposing we begin with intuitive examples of the kinds of virtues it is natural to call epistemic—traits like open-mindedness and rigor—should we conceive epistemic virtue along internalist or externalist lines? It might seem that the more voluntarist we are about the control of belief, the more likely we are to be aretaic internalists, holding that epistemic virtue is a matter of having suitably deep tendencies toward good cognitive behavior, where goodness here is a matter of adequate responding to internally accessible grounds as a basis for engaging in that behavior. The less voluntarist we are, the more likely we are to be aretaic externalists, holding that epistemic virtue is a matter of having suitably deep tendencies to form beliefs on a reliable basis.[11] But the issue is not this simple. It is true that a voluntarist may naturally construe justification and the normative aspect of knowledge as determined by how well the agent applies internally accessible standards for the kind of conduct bearing on belief formation; but a proponent of certain kinds of voluntarism could also appraise belief in relation to the conformity of the relevant conduct to external standards. It is also true that it is natural for externalists to be non-voluntarists. But as just indicated, the question of the kind of control we have over belief formation—even the kind of direct control, if we have that—is largely independent of the kind of standard we use in determining criteria for justification and knowledge. For this reason among others, I want to set aside the voluntarism issue here.

Quite apart from the question of voluntarism, an internalist view of epistemic virtue is supported by the plausibility of considering agents epistemically virtuous when, despite their believing as many falsehoods as is possible for them, they have character traits that always lead them to believe what is reasonable from the point of view of their experience and of everything they have internal access to, i.e., roughly, access by introspection or reflection. Supporting an externalist position is the plausibility of ascribing epistemic virtue to agents on the ground that, in a way that is both character-based and non-accidentally successful, they arrive at the truth in a very large proportion of the cases in which they form beliefs.

It is useful here to reflect on the notion expressed by the words "by virtue of"—which, as in the case of "dormitive virtue"—seem to mean something like "through the power of." Perhaps a virtue is a suitably deep, appropriately strong power to achieve the relevant telos. In this case, the telos is (roughly speaking) believing truths, which is good, and avoiding believing falsehoods, which is bad. This would not, however, make the trait itself intrinsically good; and we do not usually use "by virtue of" unless the base property is good. Moreover, a virtue—if we are to preserve contact with the tradition in virtue ethics and indeed with our natural admiration for virtue in persons—should be intrinsically good. The sheer power to come

to believe truths and avoid believing falsehoods is one that a robot could apparently have. Supposing it could, what conclusion might we draw?

An aretaic externalist will be inclined to note that (assuming robots lack consciousness) the way a robot arrives at truths is different from the way we do, namely on the basis of sense experience, reflection, or other accessible processes. Perhaps, then, what epistemic virtue theory needs is an "internalist externalism."[12] I find this initially plausible. But it does not explain the tendency to attribute at least certain epistemic virtues even where there is a conspicuous failure on the external dimension— provided that this failure is explainable by appeal to factors the agent could not be expected to identify or even suspect. In the case of propositions about the external world, and indeed for most empirical propositions, this kind of failure is possible.

Consider the ethical analogy: Might a person not have moral virtue despite repeatedly doing harm, provided there is a suitable explanation for the mistaken beliefs that lead to it? And isn't it, as Kant emphasized, such internal factors as intentions that determine how good a person is?[13] Certainly intentions and beliefs are central for understanding moral virtue. There is surely room to argue that moral virtue is an intrinsic property of persons and does not require their actually achieving anything in the external world— expectable though that is in normal interpersonal circumstances. The point is far from self-evident, however. Could I be a just person if my decisions never made any difference in the world because, say, I was always satisfied by perfect hallucinatory evidence, supplied by a Cartesian demon, detailing my "success" in dealing with cases of distribution and retribution? One response would be that I could be a just person, but not one whose just deeds had their normal *consequences*. There might, then, be an objective standard of conduct I must meet, but it could be internal, say a matter of my decisions and volitions. I shall not try to settle this matter here.[14]

There is a similar issue in the epistemic domain. Indeed, when it comes to epistemic virtues, there may be a distinction, not exactly paralleled in the moral case, between those that are intuitively internal and those that are intuitively external. If there is such a distinction, epistemic internalists and epistemic externalists alike must account for both kinds. Consider first conscientiousness and logicality. The first almost wears on its sleeve the centrality of a requirement of applying an internally accessible standard to the best of one's ability, and being conscientious apparently does not entail forming true beliefs about the external world. Logicality is surely possible in a high degree regardless of massive deception about the external world, provided the deception permits sufficient coherence, as a gracious Cartesian demon would surely guarantee. But now consider insightfulness, perceptiveness, and wisdom. The first, and perhaps the second and third, have, to be sure, an a priori and arguably internal aspect: A variety of insights and a kind of perception are possible in a priori matters. But people who are unqualifiedly insightful or perceptive, as opposed to, say, being logically insightful and philosophically perceptive, must be right in a good deal of cases about external matters. Perhaps they must even have a significant range of external knowledge; a perceptive person must readily see a variety of things, including many concerning people. As

to wisdom, surely this virtue (in its general form) requires some significant knowledge of human psychology.

The controversy between internalism and externalism in epistemology is deep and enduring, and my aim here has been simply to show how it affects the explication of epistemic virtue. I make no attempt to settle it, and the main points so far suggested are neutral with respect to internalism and externalism. The point is that both kinds of theory must explicate both kinds of virtue, the internal and the external. If, however, it turns out that some epistemological concepts, such as justification, are internal and others, such as perception and knowledge, are external, then virtue epistemology may bear differently on them depending on whether it is developed along internalist or externalist lines or provides space for both kinds of concepts. What I want to explore in the remainder of the essay is how the notion of epistemic virtue bears on justification and knowledge, which are uncontroversially two of the central concepts epistemology seeks to understand.

III. Epistemic Virtue and Justification

Suppose a belief is virtuous, in the sense that it is held "from" epistemic virtue, i.e., grounded in an epistemically virtuous way, as where the agent, guided by appropriate critical dispositions that express a feature of intellectual character, has formed it in the normal perceptual way or from credible testimony. Is such a belief always justified? It would seem so, but here there can be a divide between internalists and (most) externalists. For the latter, it is a contingent matter whether such a belief is justified, since it is not metaphysically necessary that these belief-forming processes be externally successful (which, for most externalists, amounts to being reliable and so producing true beliefs in at least a majority of instances). In general, however, internalists and externalists can agree that *normally* virtuous belief is justified.

The converse question is more difficult. Does justified belief have to be epistemically virtuous? It would seem possible to have a justified belief when no associated trait deserving the name "virtue" is present. A generally uninsightful person can suddenly see a pattern (in a way we think of as indicating insight) and thereby be justified in a belief about it. I do not believe that we must either find some other virtue of which the insightful belief is a manifestation or rule that the person has suddenly become insightful. Indeed, we might wonder how one can develop the traits that constitute epistemic virtues without having some justified beliefs formed in the right sort of way first, so that one can develop the right dispositions. Don't we have to reinforce children who say (and believe) the right thing under conditions that justify it, such as conditions of good light relative to color discrimination, before they develop the virtue of being observant with respect to color? We reinforce their *discriminative* belief formation, not just (and indeed not mainly) their true belief formation, as where, regarding a friend telling about an escapade, they discern credibility on one point and not another, accepting the former and withholding the latter. Must we insist that children (and improving adults) do not have any justified beliefs until they have the relevant trait?

This seems doubtful, if only because it seems impossible, in ordinary human life, to reinforce the right dispositions for developing virtue unless at least some of the time the beliefs formed in manifesting them, as with ordinary responses to clear perceptions, are sufficiently well grounded to qualify as justified. And if we are to develop rigor in children and students, don't we have to correct them (and ourselves) guided by a sense of what conditions, such as perceptual or propositional evidences, justify belief? When a belief is properly based on these conditions, is it not thereby justified? These suggestions are not meant to imply that there is no way to defend the view that justified belief is, so to speak, trait-dependent; but they do make it doubtful that the view can adequately account for epistemic development.

Analogous points certainly hold for the development of moral virtues, and I am taking it that other things equal, a good theory of epistemic virtue should largely parallel a good theory of moral virtue. In both cases, it would seem that for at least many growing children, critical faculties and good moral and intellectual habits or patterns develop gradually and in such a way that it seems at best artificial to claim that virtuous action or justified belief occurs only when there is sufficiently advanced and stable development to imply the possession of a moral or epistemic virtue.

It must be granted that if an action or belief is grounded in highly unstable psychological elements, as where the person was simply lucky not to have succumbed to temptation to do evil or to believe a false contrary proposition, then even if the elements that do ground the deed or belief are of the right sort to justify it, one might argue that it is not justified but rather, say, accidentally in conformity with justifying grounds. Suppose this diagnosis would be correct. From the fact that a certain kind of *unstable* basis would undermine justification, it does not follow that justification requires the enduring stable and global basis we consider an epistemic virtue. To show that would require far more argumentation.

If justified belief is possible apart from epistemic virtue, then being grounded in epistemic virtue cannot be what *constitutes* justified belief. And if the concept of epistemic virtue cannot be understood without an independent notion of justified belief as the kind of belief such virtue tends to yield, then the former concept is at least in part derivative from the latter. This would by no means imply that the former concept is not significant. I shall return to this matter shortly. Let us first ask the same kinds of questions about the relation between epistemic virtue and the kind of true belief that constitutes knowledge.

IV. Epistemic Virtue and Knowledge

May we say that what it is natural to call an epistemically virtuous belief that is true constitutes knowledge? Surely not, at least if expressions of virtue are taken to be praiseworthy. The same kinds of justified true beliefs that fail to constitute knowledge can be epistemically virtuous. Suppose one has a justified true belief that it is 9:15 on the basis of reading this on a normally accurate clock which, however, happens to be malfunctioning but is running in a normal-seeming way and happens to say 9:15. It looks as if a virtuous true belief will constitute knowledge only in the ab-

sence of certain defeaters, some of which (such as the present unreliability of the clock) are external to the mind of the subject.

As to the question whether knowledge, as such and not restricted to a particular kind, must be grounded in epistemic virtue, the answer is, as for the case of justification, apparently negative, and for some of the same reasons. The development of knowledge seems to precede that of epistemic virtue, and the concept of knowledge seems prior at least to the epistemic virtue concepts that are knowledge-entailing.[15] Insight may be one of these, at least if an insight cannot be false and cannot be constituted by a merely justified true belief. Even supposing that it can be so constituted—which I think is so in special cases—*insightfulness* as a trait cannot fail to yield knowledge all of the time (it may not be able to do so even in an unrestrictedly large proportion of cases, fewer than all of them). This point may be clearer in the case of wisdom: wise people not only get important things right a substantial portion of the time when significant judgment is called for; they usually know the truth of what, in such matters, they judge to be so.

We can now raise a larger question about epistemic virtue in general. Could epistemically virtuous persons have such virtue in an overall sense if they possessed only what I have called the *internal virtues*—those one can have even when maximally deceived in what one believes on the basis of their exercise—or must one have a certain proportion of the *external virtues* as well, those which, like wisdom, imply achieving knowledge, or at least achieving it when one forms beliefs on certain important matters, such as how political leaders should govern if they are to command the respect of their people? Since the notion of an epistemically virtuous person is a creation of epistemological theorizing, the best way to answer this is probably to indicate what kind of thing is at stake if we go one way rather than the other. Let me develop this suggestion.

If we try to imagine an epistemically virtuous person with only internal virtues, we are not as limited as one might think. It is true that insight and wisdom will not be possible as usually understood, but if the person is (to take some examples of important internal virtues) sufficiently judicious, imaginative, rigorous, curious, and conscientious, the errors that remain will tend to be excusable and—so far as internalism in epistemology is sound—will also tend not to undermine justification. Indeed, given how wide-ranging these virtues are, there is no question that the person would be an intellectually responsible agent. This would show that there is, if not an overarching aretaic concept, then at least a global normative notion, such as that of epistemic responsibility, which applies to people on the basis of internal virtues alone.

However, insofar as one is thinking of virtue as a kind of power to yield knowledge, and insofar as a priori knowledge and other sorts of internally grounded knowledge do not seem sufficiently important, for instance in survival or in developing scientific theories, one will deny that a person could have epistemic virtue in a full-blooded sense without external virtues. Suppose that denial is correct. It carries a price. The price would be quite high if we could unqualifiedly say that since knowledge that someone has these external virtues is not, even in one's own case, possible on the basis of internal grounds, one cannot know on such a basis even that

one has epistemic virtue oneself. But at most the implication of externalism here is that we cannot know this kind of thing *with "certainty"* (or at least Cartesian certainty). To some, this may seem unfortunate: some philosophers might want to have a way of knowing with certainty that we are meeting standards of virtue at least regarding our own beliefs, which are themselves internal, without depending on what happens outside. This kind of desire can be a powerful motive; and in ethics it can lead to the view that the actions and other elements that truly manifest our virtue are, like our beliefs, internal: They are acts and dispositions of the will, above all volitions (which are a kind of act) and intentions (which are a kind of disposition).

This internalist approach is understandable as a response to skepticism: If one cannot show that there is knowledge of the world, it would be good at least to be able to know with certainty that we have intellectual virtues. Perhaps so. But why isn't the possibility of justified belief and of what might be called "probable knowledge" enough here? Indeed, in the end it may not be possible to escape some dependence on external conditions even for achieving justified belief or morally praiseworthy action. If, as I think essential, we distinguish between situational justification—justification for holding a belief—and doxastic justification, which is justification of a belief actually held, and if the latter occurs only when the belief is in some causal sense based on what justifies it, then the notion of justified belief is not entirely internal, at least on the (debatable) assumption that the causal relation in question is not entirely accessible to introspection or reflection. Analogous points hold for action. Thus, knowledge that we have justified belief, or (as Kant saw) have done a morally praiseworthy deed, depends on our grasp of a partly external relation between grounds of justification and the belief they justify, and between grounds of moral action and the deed they warrant.

The best reply an internalist can offer here, whether a virtue theorist or not, is this. First, it may be granted that the rich, causally laden normative notions in question —those of justified belief and morally praiseworthy action, as opposed to justification *for* belief or warrant *for* acting—are partly external, and to the extent that the notion of an epistemically virtuous person or a morally virtuous agent depends on them, it is *external*. Second, it may still be true that the notion of an epistemically responsible agent, and the corresponding notion of a morally upright agent, are internal. Third, even apart from this, one could surely have justification, conceived internally, for believing one has a justified belief, and is praiseworthy for doing a deed, even if one cannot know this.

Suppose these points in partial support of an internalist version of virtue epistemology are sound. There is apparently good reason to hold that an overall account of a full-blooded intellectually virtuous person must incorporate some additional externalist elements, at least insofar as that is required to account for such a person's having knowledge of other people. The extent to which a person who is a paradigm of intellectual virtue (which I take to encompass epistemic virtue) must exhibit sound judgment, wisdom, insight, and other virtues whose characteristic expression entails knowledge is considerable. Justified belief of the relevant propositions, if short of knowledge, is not sufficient. The overall notion of intellectual virtue is, in

that sense, not purely internal. This does not entail that an internalist virtue episte-
mology cannot account for justification, but it does entail that insofar as knowledge
requires an externalist account (as seems to me the case), a full-scale virtue episte-
mology will need an externalist component.

V. The Scope and Significance of Virtue Epistemology

If I have been right in my main points about epistemic virtue, it is parallel to moral
virtue in many important ways. The notion certainly has some of the same advan-
tages, such as enabling us to describe people and to explain some of their cognitive
behavior, in terms of broad positive elements of intellectual character. We can also
state in summary fashion some of the important ideals of intellectual conduct:
Open-mindedness and rigor and insightfulness are traits covering a huge range of
territory, in terms of subject matter, cognitive style, and other variables. These intel-
lectual virtues can, to a high degree, function as autonomous critical notions in epis-
temology. Epistemological theories, whether or not they are virtue theories, can help
us understand them, but it is important that these notions be studied in their own
right (something so far too rarely done in adequate detail by epistemologists[16]). Both
epistemic and moral virtue, moreover, have internal and external elements and, as
that suggests, can be given (plausible) internalist or externalist accounts.

On strong interpretations of virtue ethics, the notion of a virtue is conceptually
more basic than that of a morally right action (or any other "individual" notion). A
virtue epistemology can be given a similarly strong interpretation on which, above
all, a justified belief is one grounded in an epistemic virtue, where the latter concept
is more basic. I have suggested that epistemic virtues are not conceptually in-
dependent of particular cases of justified belief (or of rational belief or some other
kind of individually characterized belief). It seems to me, indeed, that there is less
difficulty — even if not ultimate success — in explicating epistemic virtue as a kind
of trait that yields justified beliefs and, in some cases, knowledge, than in explicating
justified beliefs or knowledge by appeal to an independent notion of virtue and con-
struing them as the kind of belief it tends to produce.

This is not to imply that no theory plausibly called a virtue epistemology can
avoid circularity, especially if the theory concerns knowledge rather than justifi-
cation. Even if the easiest and perhaps clearest way to explicate epistemic virtue, as
relevant to knowledge, is to presuppose that we know things of certain kinds in cer-
tain ways — say through perception and intuition — and then characterize epistemic
virtue as an internalization of standards or rules for the proper uses of these facul-
ties, one need not appeal to the concept of knowledge itself in defining such virtue,
nor is it clear that the relevant standards are fully understandable apart from a con-
cept of virtue. But one will surely have to presuppose a way of telling when the cog-
nitions that express a virtue are *true*. Here, I think, we will have to trust what might
be called the standard basic sources of knowledge: perception, consciousness, mem-

ory, and reason.[17] A similar point might hold for justification. A moderate virtue epistemology can grant such a dependence of basic epistemological concepts on basic sources (if not these, then others with a similar role). However this turns out, we are not forced to adopt one stark theoretical option or the other in epistemology, being either strong virtue theorists or treating epistemic virtue as a mere adjunct to other notions. The notions of epistemic virtue and of justification are so closely related that a full-scale account of either one can be used as a basis for checking on and even extending any account of the other.

The possibility of explicating epistemic virtue in terms of non-aretaic standards of justification and knowledge does not imply the reducibility of virtue concepts to other kinds, such as rule concepts. That rigor and insightfulness and logicality, for instance, can be explicated by appeal to such notions as enduring success in forming justified beliefs, and acquiring knowledge, in response to certain kinds of grounds does not entail that this is all there is to those virtue notions. Virtue concepts have a certain descriptive autonomy. They may be fully understandable only if such response tendencies are properly appreciated, but they have an open-endedness that is at least not easily captured in non-aretaic terms.

Moreover, even if we should be able to account for the notions of knowledge and justified belief by appeal to non-aretaic concepts, the notion of intellectual worth, like its counterpart notion, moral worth, cannot be understood apart from virtue notions. A person's intellectual worth is in part a matter of epistemic character. No matter how many justified beliefs a Cartesian demon might give one, intellectual worth would not be implied; nor would a wide knowledge, arrived at in an inappropriate way, count toward one's having such virtue. Just as actions must be rooted in one's character to count toward one's being a morally good person, beliefs must be so rooted to count toward one's having good intellectual character.

The descriptive autonomy and normative power of virtue concepts suggests a further point, one that has not to my knowledge been noted. In ethics, that autonomy and power may in part account for W. D. Ross's plausible and influential view that practical wisdom (which is largely a matter of intellectual and moral virtues) is needed to deal with conflicts of prima facie duties.[18] Let me suggest a similarly important role for epistemic virtues.

Even if we suppose that the basic grounds of justification, and indeed the most basic ingredients of the notion of justification, are non-aretaic, epistemic virtue concepts can play an important epistemological role. Consider a case in which we have conflicting evidence, say where witnesses disagree on the details of an accident. On the basis of their testimony and one's own memory as well, one might have prima facie justification for each of two incompatible hypotheses about the cause of the accident. If one must make a determination, it may be helpful to think in terms of people who are role models of judicious judgment or of what a rigorous appraisal of the evidence by such a person would call for. It could turn out in such cases that even if one's prima facie justification depends entirely on evidential grounds of the usual sort, one's *overall* justification for believing what one does depends in part on how well one realizes the relevant virtue. Roughly, the idea is that even prima facie justification for believing, say, that icy road conditions caused the crash, depends on per-

ception, testimony, and inference, one's overall justification for believing this as opposed to a competing view is its support from a sense of overall plausibility rooted in one's judicious epistemic character. In practice this kind of role for epistemic virtue can be a vitally important matter.

This point leads to a further one. Regardless of which notion is more basic, if either of them, a justified belief should be of a *kind* that arises or at least can arise from an epistemic virtue, and an epistemic virtue should be the *kind* of trait whose most characteristic expression is the formation of a justified belief; and similarly for knowledge in relation to the external virtues. The notion of epistemic virtue can, then, serve as a basis for developing illuminating necessary and sufficient conditions for justified belief and knowledge, even if these concepts are not analyzed by appeal to it as the more basic notion.

A parallel point holds for epistemic deontology of the most plausible kind. A justified belief should be in some sense epistemically responsible, and an epistemically responsible agent should most characteristically form justified rather than unjustified beliefs. Again, we might develop illuminating necessary and sufficient conditions for justified belief by appeal to epistemic responsibility—which, indeed, can be conceived as a kind of master epistemic virtue—even if the notion of such responsibility is not a basis for an analysis of the concept of justification. (The intuitive appeal of an epistemic responsibility approach seems to me greater for explicating justification than for explicating knowledge, but I cannot pursue this here.) Like the individual epistemic virtues, however, the notion of individual epistemic responsibilities—such as observing a scene more closely and checking one's reasoning on a problem—can be incalculably useful in guiding both the process of education and the exercise of judgment. There are epistemic virtues; they have explanatory interest regarding cognition, as well as a normative role in daily life and an analytic role in epistemology. There are also epistemic duties, such as the duty to avoid hasty judgment and invalid inference. These are regularly fulfilled by epistemically virtuous agents.

There is yet another point important for appreciating both the analogy between virtue epistemology and virtue ethics and the significance of each. In epistemology as in ethics, the exercise of virtue takes one beyond doing one's strict duty. This is another valuable feature of virtue epistemology: it focuses attention not just on what justification and knowledge are, but also on what counts as admirable, as opposed to merely acceptable, pursuit of them. Epistemic virtue, being in part a matter of actively seeking to learn, produces much more than a body of generally justified beliefs and diverse knowledge; it yields, in those with moderate good fortune, a measure of unity in their view of the world.[19]

Notes

1. For recent theories constituting one or another kind of virtue epistemology see, e.g., Ernest Sosa, *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991); Jonathan Kvanvig, *The Intellectual Virtues and the Life of the Mind* (Lanham: Rowman and Littlefield, 1992); James A. Montmarquet, *Epistemic Virtue and Doxastic Responsibility* (Lanham: Rowman and Littlefield, 1993); Linda Zagzebski, *Virtues of the Mind* (Cambridge:

Cambridge University Press, 1996); and John Greco, "Virtues and Vices of Virtue Episte-mology," *Canadian Journal of Philosophy* 23 (1993): 413–32 and "Agent Reliabilism," *Philosophical Perspectives* 13 (1999). A instructive position for comparison is Alvin Plantinga's proper functionalism; see especially *Warrant and Proper Function* (New York: Oxford University Press, 1993). For a wide-ranging valuable treatment of the recent literature see Guy Axtell, "Recent Work on Virtue Epistemology," *American Philosophical Quarterly* 34 (1997): 1–26.

2. I have explicated the notion in question in "Acting from Virtue," *Mind* (1994), reprinted in my *Moral Knowledge and Ethical Character* (New York: Oxford University Press, 1997).

3. In "Internalism and Externalism in Moral Epistemology" in my *Moral Knowledge*, I have critically assessed this kind of conception as an example of an epistemically externalist view and indicated how Kantian and other kinds of theory offer an epistemically internalist conception of moral virtues and other moral notions.

4. This point is perhaps implicit in Aristotle's *Nicomachean Ethics* but seems clearly intended by W. D. Ross in *The Right and the Good* (Oxford: Oxford University Press, 1930), ch. 2.

5. One reason the analogy is rough is that whereas two obviously incompatible actions can be equally well justified morally for an agent and well enough justified for one to be obligatory, two obviously incompatible propositions almost certainly cannot be thus justified for a person: If I have such good justification for *p* that I "must" believe it, I almost certainly do not have this degree of justification for an obviously incompatible proposition, such as not-*p*.

6. Thus, we find Linda Zagzebski saying, "Knowledge is a state of belief arising from acts of intellectual virtue." See her precis of *Virtues of the Mind, Philosophy and Phenomenological Research* 60, 1 (2000): 176. (The symposium that follows, with essays by William Alston, John Greco, Hilary Kornblith, Jonathan Kvanvig, and Amelie Rorty, with replies by Zagzebski, is quite pertinent to this essay.)

7. One might think that no one regards belief as action, but at least one writer on epistemic obligation says, "Sometimes it makes sense to treat beliefs like other actions and to evaluate their practical or prudential merit." See Richard Feldman, "Epistemic Obligations," *Philosophical Perspectives* 2 (1988), sect. 1. For some reasons to avoid construing belief as action and an appraisal of various forms of doxastic voluntarism, see my "Doxastic Voluntarism and the Ethics of Belief," *Facta Philosophica* 1 (1999): 1.

8. For discussion of the nature and appraisal of various kinds of failure to respond adequately to evidence, see Christopher Hookway's "Epistemic Akrasia and Epistemic Virtue," in this volume.

9. This is probably not uncontroversial; for a certain kind of subjectivist approach to rational action the motivating belief(s) need simply have the right kind of content, e.g., that the action in question is the best way to achieve the relevant goal, and need not themselves be rational. I cannot recount difficulties for this view here, but surely it would be at best a strange irony if a virtue epistemology took this kind of view, on which the basis of a rational action in the agent can be so heavily dependent on a belief that is intuitively without epistemic virtue. One might also note that given a sufficiently strong desire to believe only truths (or only truths apart from special exceptions), together with a guiding belief that in general this requires believing on the basis of evidence, we might exercise rational control over an ability to believe propositions "at will." This is too large an issue to pursue here. Some of the relevant considerations are explored in my "Doxastic Voluntarism," cited above.

10. I take the intrinsically good to be good in itself and the *inherently good* to be such that an appropriate experience of it is intrinsically good. I prefer to consider only experi-

ences intrinsically good, but I take indefinitely many kinds of things to be inherently good and hence sources of non-instrumental reasons for actions (such as contemplating the items in question). These notions are explicated (with references to relevant literature) in chapter 11 of *Moral Knowledge*. In this essay I ignore the experientialist view of intrinsic goodness and construe virtues (which are of course not experiences) as intrinsically good even though I prefer to consider them inherent goods.

11. This is not to imply that an aretaic internalism must be an epistemic deontologism. Despite the common association between internalism and deontologism, even a strong accessibility internalism need not be deontological.

12. See William P. Alston, "An Internalism Externalism," in his *Epistemic Justification* (Ithaca: Cornell University Press, 1989).

13. See the *Groundwork*, sects. 395–96.

14. Just how much objective success is implied in a virtue concept is discussed in the symposium on Zagzebski's *Virtues of the Mind* cited above. This is explored for moral virtues in my "Acting from Virtue" (in *Moral Knowledge*), where I argue that there is apparently some kind of limit, variable across virtues and circumstances, to how far wrong one can go and still retain the virtue in question. A related idea is that a virtue is a skill, as argued by Paul Bloomfield in "Virtue Epistemology and the Epistemology of Virtue," *Philosophy and Phenomenological Research* LX, 1 (2000): 23–43; but I would say that (as "skill" is used in English) virtues entail skills but are not equivalent to them.

15. Cf. Ernest Sosa's comment that "the subject bootstraps up from animal to reflective knowledge" and his point that human reflective knowledge is most likely to depend ultimately on unreflective knowledge, since we cannot climb infinite ladders of reflection." See "Intellectual Virtue in Perspective," in *Knowledge in Perspective,* pp. 284 and 290 respectively.

16. Here Zagzebski, *Virtues of the Mind* (especially in Part II), has offered some of the more detailed explorations available of representative intellectual virtues.

17. I think this point is consistent with the subtle treatment of the circularity problem by Sosa, *Knowledge in Perspective*. For a related approach to explicating the epistemic virtues without vicious circularity, see Zagzebski's *Virtues of the Mind*.

18. See *The Right and the Good,* ch. 2.

19. This essay benefited greatly from discussions at the Santa Barbara Conference on Epistemic Virtue and Duty in 1999 and at Wake Forest University in the same year. I also want to thank John Greco, Christopher Kulp, and especially Michael DePaul, my commentator at the conference, for detailed helpful comments.

7

THIN CONCEPTS TO THE RESCUE

*Thinning the Concepts of Epistemic Justification
and Intellectual Virtue*

Heather D. Battaly

Much of contemporary analytic epistemology has been devoted to an-
alyzing the concept of epistemic justification and providing neces-
sary and sufficient conditions for its application. But, despite our best efforts, there
has been persistent and widespread disagreement about these conditions, much of
which appears to be irresolvable. Indeed, rival analyses have fueled the now notori-
ous debate between internalists, who place accessibility requirements on justifi-
cation, and externalists, who reject such requirements. At best, our analysis of justi-
fication has led to a stalemate, at worst to conceptual turmoil.

The fruitlessness of the analysis of justification has led both William Alston
and Stewart Cohen to argue that epistemologists would do well to abandon it and
pursue more worthwhile projects in its place.[1] Interestingly, the very same frus-
tration has provided part of the impetus for the recent turn to virtue epistemology.
The defining feature of virtue epistemology is its focus on the intellectual virtues
and vices of agents instead of justification, knowledge, or any other evaluation of
belief. Virtue epistemologies will either eliminate the concept of justification or
define it in terms of intellectual virtue, thus relying on our familiarity with the
concept of virtue to clarify the concept of justification. So, part of its appeal is its
supposed ability to circumvent the irresolvable debates about justification. But,
ironically, competing analyses of intellectual virtue indicate that virtue epistemol-
ogy is poised for its own irresolvable debates. According to James Montmarquet,
intellectual virtues are characterized by a motivation for truth, but need not be
truth-conducive.[2] For Linda Zagzebski, they involve both a motivation for truth
and reliable success in attaining it; while for Ernest Sosa and Alvin Goldman, they
are reliable faculties that need not involve motivations.[3] Echoing the advice of Al-
ston and Cohen, must we then abandon the analysis of intellectual virtue and
move on to yet another project? Is epistemological analysis destined to become in-
fected with fruitless debates and subsequent calls for abandonment? Thankfully
not, or so I will argue.

I will show that we can circumvent such debates by recognizing that the con-

cepts of justification and intellectual virtue are thin. Each is thin because it has multiple conditions of application; different speakers appeal to different combinations of these conditions in applying it; and there is no definite answer as to which of these combinations is necessary, or which is sufficient, for its application. Pace Alston and Cohen, I will argue that both our ordinary and technical concepts of justification are thin. It is for this reason that they do not support a meaningful debate between internalists and externalists. Had we recognized this earlier, we could have avoided the debate entirely simply by precisifying our concepts of epistemic justification. Fortunately, we can still prevent virtue epistemologists from picking this poisonous apple. I will argue that because our ordinary and technical concepts of intellectual virtue are thin, virtue epistemologists can inoculate themselves against a misguided debate about virtue by recognizing that they are thickening these thin concepts of virtue in different ways. To preclude a fruitless debate about virtue they need only acknowledge that they are analyzing different thick concepts. Thanks to thin concepts, neither the analysis of epistemic justification nor the analysis of intellectual virtue need be abandoned.[4]

I. Alston on Epistemic Justification

Alston argues that much of the contemporary debate over epistemic justification is misdirected. In his words, "a large proportion of contemporary epistemologists . . . have been misguided in their researches, fighting under a false banner, engaged in a quixotic tilting at windmills."[5] He sometimes writes as if this debate were about the *concept* of epistemic justification, and at other times implies that it concerns the *referent* of "epistemic justification"—the property to which the term refers. Here, I concentrate on the former interpretation.

According to Alston, the parties to the debate disagree about the necessary conditions of justification, but agree, for each proposed condition, that its satisfaction is desirable vis-à-vis the basic aims of cognition. What conditions do the parties propose? And, how, given the ever-expanding mess of claims about justification in the recent and current literature, do we even begin to identify them? Alston suggests a basic framework for the dispute that allows us to enumerate several of the proposed necessary conditions. He assumes that for all concerned parties, the justification of S's belief that p has something to do with the relation of that belief to S's grounds for it. For Alston, S's grounds are constituted by her experiences and/or her prior justified beliefs or knowledge. Given this framework, he identifies nine proposed necessary conditions of justification:

(1) having a ground which *is* sufficiently indicative of the truth of one's belief;
(2) basing one's belief on a ground which *is* sufficiently indicative of the truth of that belief;
(3) basing one's belief on a ground which one would on adequate reflection *believe* to be sufficiently indicative of its truth;
(4) having cognitive access to the ground of one's belief—e.g., being able to spot the ground of one's belief on reflection;

(5) having cognitive access to the adequacy of that ground — e.g., believing on adequate reflection that one's ground is sufficiently indicative of the truth of one's belief;

(6) knowing, or justifiably believing, that one's ground is indicative of its truth;

(7) satisfying one's intellectual obligations;

(8) acquiring one's belief from the exercise of an intellectual virtue; and

(9) having a coherent system of belief with which the belief in question coheres.[6]

According to Alston, internalists who, in rejecting (2), deny that reliability is necessary for justification do *not* deny that it is epistemically a good thing to form beliefs in such a way that they are thereby likely to be true. Nor do externalists, who abjure accessibility requirements like (3) and (4), denounce the epistemic value of having access to one's grounds and their adequacy. In fact, though internalists and externalists disagree about which of (1)–(9) are necessary for justification, they agree, for each of (1)–(9), that the states that result from satisfying it are desirable given the basic aims of inquiry.[7] In other words, they concur that all of these states are epistemic desiderata.

Alston suggests that there are certain features of the debate over justification that are best explained by supposing that there is no unique concept about which the parties are disagreeing. He claims that the term "justification" was originally imported into epistemology from a discussion of the justification of voluntary action. An act was said to be justified just as long as it was permitted by the relevant legal, moral, or institutional norms. Analogously, *S*'s belief that *p* was said to be epistemically justified just as long as it was permissible for *S* to believe that *p*; i.e., *S* was violating no epistemic obligations in so doing. So, our original interpretation of "epistemic justification" was a deontological one. Alston rejects this conception on the grounds that it entails direct voluntary control over belief, and he maintains that our abandonment of this, the most natural interpretation of "justification," really ought to be accompanied by our elimination of the term from epistemology. But, and here lies the rub, it would be nearly impossible to eliminate the term because of its ubiquity in contemporary epistemological discourse, *even in "nondeontological circles."*[8] According to Alston, "there are other roots of the epistemological use of the term . . . and whether or not these roots can nourish "justified" with linguistic propriety, they unquestionably do influence current epistemological thinking about what is called 'justification.'"[9]

To bolster his case, Alston responds to the objection that there *is* a theoretically neutral way of identifying what epistemic justification is, and thus, that there is some common concept about which the parties disagree. His response has two parts. He begins by arguing that since we cannot find any theoretically neutral way to *uniquely* identify the shared concept about which the parties purportedly disagree, it is reasonable to conclude that there is no such concept. To that end, he argues that we cannot use any of the proposed necessary conditions, (1)–(9), to identify a shared concept of justification because for each of these conditions, some of the parties will deny that it is necessary. Nor can we use the descriptions "cog-

nitive state that is desirable from the point of view of maximizing true belief and minimizing false belief" or "that which, in the absence of Gettier problems, converts true belief into knowledge." The former does not uniquely identify justification, and the latter is not theoretically neutral. The former also applies to true belief, and true belief is distinct from justified belief. The latter is not theoretically neutral because if it turned out that reliable belief formation converted true belief into knowledge, internalists would recant the claim that justification converts true belief into knowledge.

Alston then considers the possibility that the parties share a concept that cannot be uniquely identified by a theoretically neutral definite description—a concept for which they can formulate no formal definition. In his words, "there are other ways of zeroing in on a topic than by giving a definite description that uniquely picks out the target."[10] For example, one might feasibly "zero in" on justification by deploying shared paradigms of justified and unjustified beliefs in much the same way that one might "zero in" on the subject matter of our ordinary concept of dog by deploying shared paradigms of dogs and non-dogs. To explicate,

> Even if we can't provide a general formula for saying what justification is without getting into theoretically controversial matters . . . we can trot out some . . . uncontroversial paradigms of justified and unjustified belief . . . that will enable any sufficiently alert spectator to form . . . the concept of justification shared by all the parties to the disputes, even if no one can give the concept an informative pre-theoretical *definition*.[11]

So, even if we can't spell out theoretically neutral necessary and sufficient conditions for justification, we can still grasp the concept that the parties purportedly share by studying paradigm cases of justified belief and appropriate ways of extrapolating from them. Alston argues that this picture, though feasible, cannot be ascribed to the parties in question because paradigm cases of justified and unjustified beliefs, and ways of extrapolating from them to other cases, differ from one party to the next.

Alston concludes that the parties involved in the most heated debates over justification are working with different concepts. Consequently, their debates have been and will continue to be misdirected. For, if there is no single concept of justification about which they disagree, it is misguided to think that their debates will eventually determine which of them is correct about justification. Interestingly, Alston maintains that those who share the same concept of justification, e.g., reliability, can substantively debate less fundamental issues, e.g., whether a reliable-indicator or a reliable-process account is preferable. Nevertheless, he ultimately recommends that we forsake the analysis of justification for an investigation of the epistemic desiderata that can be salvaged from it.[12]

II. Cohen on Epistemic Justification

Cohen argues for the unambiguous conclusion that we should abandon the conceptual analysis of epistemic justification. He enumerates three projects that epistemologists who analyze the concept of justification might feasibly be pursuing. They might be (i) analyzing a concept that is expressed by a term regularly used in ordinary discourse; (ii) analyzing a concept that is expressed by a term rarely used in ordinary discourse; or (iii) treating "justified belief" as a technical term. He argues that the first project, though philosophically significant, cannot truly be attributed to these philosophers, and that the second and third projects garner little philosophical worth. Since there is no philosophically significant issue about justified belief, we would do well to forsake the conceptual analysis of justification for the philosophically rich areas of epistemology: theories of knowledge, evidence, reliability, and rationality.[13] Here, I focus on the first and third projects.

Cohen believes that it *is* philosophically worthwhile to analyze concepts that play an important role in the way we think about ourselves. On his view, such concepts are expressed by terms regularly used in ordinary discourse. For example, because we, in ordinary discourse, regularly characterize ourselves as knowing things—knowing that the Rendezvous serves pork ribs, knowing that the new incinerator *will* reduce unemployment but increase air pollution—it *will* be worthwhile to determine exactly what we mean by the word "know." That is, it will be worthwhile to analyze the concept expressed by "know." According to Cohen, theories of knowledge are just analyses of the concept that we (in ordinary discourse) express by our regularly used term "know." As such, they can and should account for our everyday intuitions about knowledge. In contrast, theories of justification cannot account for our everyday intuitions about justified belief because, outside of philosophy, we have no such intuitions and no such regularly used term "justified belief." In Cohen's words, "there is no ordinarily used expression 'justified belief.' The description of beliefs as justified is something that occurs primarily in conversations among working epistemologists."[14] Though we frequently use "reasonable," "rational," and "reliable" to assess beliefs, we use "justified" only in connection with voluntary action. Consequently, "we cannot say that the project of the justification theorist is to provide an analysis of the concept that we ordinarily express by 'justified belief.'"[15] Moreover, since "justified belief" is not a phrase we regularly use in ordinary discourse, justification is not a concept that is central to the way we think about ourselves. If its analysis is philosophically significant, it must be so for some other reason.

Cohen suggests that the parties may instead be treating "justified belief" as a technical term. We do not define technical terms by reflecting on how they would be used in ordinary discourse but create them to fulfill specific theoretical purposes and stipulate their meanings accordingly. Cohen remarks that where there is agreement that a technical term is to play a specific theoretical role, there can be a substantive philosophical debate about which of several competing definitions of the term best plays that role. On his view, most of the disputants agree that the term "justified belief" applies to many of our everyday empirical beliefs and all be-

liefs that count as instances of knowledge, while having something to do with the goal of attaining truths and avoiding falsehoods. He argues that these meager constraints on the theoretical role of "justified belief" "are not sufficient to adjudicate the major controversies that dominate the literature."[16] They cannot support a meaningful debate between internalists and externalists. To illustrate, let's elect (a) "belief that is based on a ground which is sufficiently indicative of its truth," the representative of externalist definitions; and (b) "belief that is based on a ground which one would, on adequate reflection, believe to be sufficiently indicative of its truth," the representative of internalist definitions. If the theoretical role of "justified belief" is simply that of applying to most of our empirical beliefs and all instances of knowledge, while having something to do with the goal of believing truths, these definitions are equally satisfactory. Each is consistent with the aforementioned constraints in part because "the cognitive goal of believing all and only truths admits of both an internalist and an externalist interpretation."[17] So, if "justified belief" just is a technical term that is meant to play the aforementioned theoretical role, it will be misguided to claim that (a) but not (b) expresses *the real* concept of justification (or vice versa).

III. Thin Concepts

Though I agree that the debate over justification is misdirected, I do not find either of the aforementioned arguments completely successful. In my view, Alston's conclusion is premature because he does not seriously consider the possibility that the parties share what I will call a "thin" concept of justification. And, while Cohen does allow for a thin *technical* concept of justification, he does not consider the possibility that we can use *ordinary* phrases other than "justified belief" to express a thin concept of justification.

To begin explaining what I mean by "thin concept," I borrow an example from one of Alston's earlier works, *Philosophy of Language*.[18] There, he argued that the term "religion" is subject to combinatorial vagueness. By this, he meant that the term has several independent conditions of application and that there are no sharp distinctions between "those combinations of conditions which are, and those which are not, sufficient and/or necessary for [its] application."[19] To explicate, he argued that when we examine paradigm cases of religion, we find that they exhibit certain features, each of which "seems to have something to do with making them religions."[20] These features include (i) a belief in supernatural beings, (ii) a moral code thought to be sanctioned by them, (iii) the identification of sacred objects, (iv) ritualistic behavior, (v) prayer, (vi) a world view that specifies the role of the individual in the world as a whole, (vii) the individual's adherence to that world view, and (viii) a social organization founded on the aforementioned features.[21] When all of these conditions are satisfied, as in the case of Orthodox Judaism, fluent speakers will agree that "religion" clearly applies; and when none are satisfied, as in the case of gardening, we will agree that it clearly does not apply. But, there will be cases in which some of the conditions are satis-

fied and others are not, where we will be uncertain, or will disagree, about whether it applies.

For instance, what are we to say about Communism which satisfies (iii), (iv), (vi), (vii), and (viii), but neither (i), (ii), nor (v)? According to Alston's *Philosophy of Language*, our inability to definitively say whether Communism is, or is not, a religion is due not to a lack of information about Communism, but to the meaning of the term "religion." "Religion" gets its meaning by being applied to paradigm cases and is then "extended to other cases that do not differ from the paradigm in too many respects," but "it is *impossible* to say exactly how many respects are too many" or which conditions are the most important.[22] In other words, there is no definite answer to the question of what combination of these seemingly relevant conditions is necessary for the application of "religion," or of what subset, short of the whole, is sufficient. There simply is no sharp distinction between the combinations of conditions that are, and those that are not, necessary or sufficient for its application. Since the meaning of "religion" does not determine which conditions we should use to extend its application, selecting any one subset of conditions as necessary and/or sufficient [e.g., (i)–(iv)] is just as arbitrary as selecting any other [e.g., (vi)–(vii)]. The only way to resolve our uncertainty about the application of "religion" and the disagreements between those who have selected different combinations of conditions is to make its meaning more precise — to stipulate which conditions are to count as necessary and sufficient for its application.

Using Alston's account of "religion" as a starting point, we can say that the concept of religion is thin because (1) it has multiple conditions of application; (2) different fluent speakers appeal to different combinations of these conditions (C_1, C_2, C_3 . . .) in applying it; and (3) there is no definite answer as to which of these combinations (C_1, C_2, C_3 . . .) is necessary, or which is sufficient for, its application. This formulation of a thin concept allows fluent speakers to agree on some of the conditions of the application of that concept; i.e., it allows some of the conditions of its application to be fixed. To illustrate, we still employ a thin concept of religion when we agree that believing in supernatural beings is a necessary condition of religion, but disagree about the relevance of conditions (ii)–(viii). For though we concur that religion requires a belief in supernatural beings, different fluent speakers will still extend the concept of religion in different directions, and there will still be no definite answer as to how the concept should or should not be extended. In other words, our concept of religion will still be thin because even though (i) is fixed as a necessary condition of its application, there will be no definite answer as to which combinations of (ii)–(viii) are necessary or sufficient.

Thinness is thus a matter of degree. The more conditions on which fluent speakers agree — the more conditions that are fixed — the less thin our concept. Were we to agree that conditions (i) *and* (ii) are necessary, we would be employing a slightly thicker concept of religion. And, if we were to recognize that (i)–(viii) seem to have something to do with religion without agreeing on *any* combination of them (short of the whole) that is sufficient or necessary (as Alston claimed above), we would be employing a thinner concept of religion. In fact, we would be

employing a *maximally thin concept* of religion. When a concept is maximally thin, fluent speakers will have enumerated several seemingly relevant conditions of its application, but will not have agreed on *any* combination of them (short of the whole) that is sufficient or necessary for its application. Maximally thin concepts can (but need not) be family resemblance concepts. Accordingly, there may be no single condition that is common to all of our applications of the concept of religion—there may be no single feature that all religions share. Instead, religions will resemble one another as members of a family do—by means of a complicated network of similarities, some overlapping and some criss-crossing.[23]

Thin concepts should be contrasted with thick concepts. Thickness is also a matter of degree. A concept is maximally thick when all of the necessary and sufficient conditions for its application are fixed—when all of its boundaries are precise. We can think of maximally thick concepts as Fregean concepts. According to Frege, the concept of *P* must identify what is distinctive about *P*—what separates *P* from *everything* else. In his words, "A definition of a concept . . . must be complete; it must unambiguously determine, as regards any object, whether or not it falls under the concept . . ."[24] Thus, the definition of the concept of prime must map not only every number to a truth-value, but every non-numerical object (e.g., the Moon) to a truth-value.[25] If the definition fails to do so, it fails to specify a sharp boundary and consequently fails to specify a concept. For Frege, "To a concept without sharp boundary there would correspond an area that . . . faded away into the background. This would not really be an area at all; and likewise a concept that is not sharply defined is wrongly termed a concept."[26] It may be unrealistic to think that we can make our thin concepts maximally thick or that doing so would always be helpful. But, to the extent that we are engaged in philosophical analysis, it will often be useful to make our thin concepts thick*er*. We can thicken the thin concept of religion by stipulating that (i) and (iii) are the necessary and sufficient conditions for its application, even if (i) and (iii) are themselves thin and we cannot sharpen all of their boundaries.

Many of our concepts are thin in the way that I have described: among them are our concepts of science, art, literature, sport, and, as I will argue, epistemic justification and intellectual virtue. To borrow a useful metaphor from Michael Lynch's *Truth in Context*, a thin concept is like a roughly drawn sketch that can be completed in different ways.[27] How one completes the sketch will depend on how one thickens the thin concept. Different ways of thickening the concept can result in drastically different pictures. But, no one way of completing the sketch is any more correct or less arbitrary than any other. Since there is no single right way to fill it in, it will be misguided to argue about which of them is the *real* picture, or the *real* concept.

Even though thin concepts and vague concepts share many of the same properties, I use "thin" rather than "vague" because, ironically, "vague" has been allotted a rather restricted meaning in the current literature. The paradigm cases of vague concepts are the concepts of bald, tall, and old, which exhibit degree vagueness rather than combinatory vagueness.[28] As such they admit of borderline cases, have fuzzy boundaries, and are susceptible to sorites paradoxes. Though thin con-

cepts also have borderline cases and fuzzy boundaries, their association with sorites paradoxes is tangential.

Perhaps my caution about using "vague" is ultimately unwarranted. After all, in their reader on vagueness, Rosanna Keefe and Peter Smith do mention concepts whose vagueness arises because of their "multi-dimensionality."[29] If we were to throw caution to the wind and understand thin concepts as vague, I suggest we do so under the rubric of semantic indecision. According to David Lewis, "vagueness [is] semantic indecision: where we speak vaguely, we have not troubled to settle which of some range of precise meanings our words are meant to express."[30] On this view, vague concepts and words can be made precise in a range of different acceptable ways. For instance, "system involving belief in supernatural beings" and "system involving ritualistic behavior" are equally good but distinct ways of sharpening "religion," and it is arbitrary to select one of them over the other. Communism is a borderline case of religion because according to some acceptable precisifications, it is a religion, and according to others, it is not. Orthodox Judaism is clearly a religion because it counts as a religion under all permissible precisifications, whereas gardening is clearly not a religion because no permissible precisification counts it as one.[31]

IV. The Thin Concept of Justification

We are now ready to evaluate the arguments of Alston and Cohen. Recall that Alston argues that we cannot find any theoretically neutral way to identify the concept of justification that the parties allegedly share. He first tries to find a theoretically neutral definite description that uniquely picks out what they call "justification." When this fails, he suggests that we look for a shared concept by studying paradigm cases of justified belief. But, since paradigm cases and ways of extrapolating from them differ from one party to the next, he concludes that there is no shared concept of justification.

Alston's first attempt to find a shared concept of justification fails to detect thin concepts altogether. He suggests that we try to locate a common concept by looking for a description that distinguishes justification from all other commendable epistemic states. In other words, he tries to find a Fregean concept—a maximally thick concept—that the parties share. But, if what they share is a thin concept of justification, we will not find it by searching for one that is maximally thick.

Do we have reason to think the parties share a thin concept of justification? Let's compare our epistemic situation to the situation in religion. Accordingly, let's begin with a clear case of justified belief. Suppose that Trudy believes, on the basis of her current visual experience of her front steps, that they are covered with snow. She sees her front steps in broad daylight and at close proximity and is neither subject to hallucinations or the whims of an evil demon, nor believes herself to be so. Not only *is* her experience sufficiently indicative of the truth of her belief, but her belief coheres with many of her other beliefs (that it was snowing this morning, that her son did not have time to shovel the front steps, etc.) Moreover, on reflec-

tion, Trudy correctly believes both that her belief is based on a ground that is sufficiently indicative of its truth, and that her belief does not violate any epistemic obligations.

When we examine clear cases of justified belief, like this one, we find that they exhibit certain features, each of which seems to have something to do with making them justified beliefs. These features include the proposed necessary conditions (1)–(9), along with several others: (10) having a belief that is good from the point of view of maximizing truth and minimizing falsity; (11) having adequate evidence for a belief; and (12) having a belief that if true, is likely to constitute knowledge. When all of these conditions are satisfied, as in Trudy's case, epistemologists agree that "justified belief" clearly applies; and when none are satisfied, we agree that it clearly does not apply. But, the main point is that there are cases in which some of the conditions are satisfied and others are not where internalists and externalists disagree about whether it applies. Suppose Bob's cognitive situation is just like Trudy's except he does not believe that his belief is based on a ground that is sufficiently indicative of its truth or that it fails to violate any epistemic obligations. As we are well aware, this disagreement does not result from a lack of information about the case—we all agree on the pertinent facts. Rather, as was the case with "religion," it results from the meaning of the term in question. The meaning of "justified belief" does not determine which combination of conditions (1)–(9) is necessary for its application, or which, short of the whole, is sufficient. Consequently, there is no definite answer as to how we should extend its application. So, if our concept of religion is thin, our concept of epistemic justification is also thin.

In his search for a maximally thick concept, Alston considers that the parties may have agreed that justification is a cognitive state that is desirable from the point of view of maximizing true belief and minimizing false belief.[32] If the parties have agreed on this, they share a thin concept of justification. For, though they agree on condition (10) above, they will still disagree about the relevance of (1)–(9). They will still fill in the sketch, now constrained by (10), in different but equally correct ways. The internalist who takes (5) to be a necessary condition of justification will deny that Bob's belief is justified, while the externalist who does not will assert that it is. We recall that Alston rejects this description because it fails to uniquely identify justification—because it fails to be maximally thick. This shows that his initial test for shared concepts admits of false negatives—it is not fitted to detect shared concepts that are thin.

Alston *is* aware of this problem. He recognizes that a group can share a concept while lacking a definite description that uniquely picks out its subject matter, and he attempts to adjust his search accordingly. But, he explores only one way to locate a thin concept of justification, which he rejects on the grounds that the parties do not share paradigms of justified belief, or at least do not use the same ways of extrapolating from those paradigms. I think that the parties do share paradigms of justified belief, but agree that they extend the application of "justified belief" differently. However, contra Alston, the fact that they do does not show that they fail to share a concept of justification. For they may still share a thin concept that they have thickened in incompatible ways. Alston concludes that the parties are not

sharing a concept of justification without even considering the possibility that they may have agreed on some aspects of the theoretical role of "justification." Following Cohen, we might construe the parties as having agreed that "justification" has something to do with the goal of attaining truths. This limited agreement on the theoretical role of "justification" indicates that the parties share a thin concept. Alston's failure to explore this possibility is made all the more odd by his earlier admission that the parties agree that justification is a cognitive state that is desirable from the point of view of maximizing true belief and minimizing false belief. Though it would have been natural for him to return to this description in his endeavor to detect thin concepts, he fails to do so.

I conclude that Alston's search for a shared concept of justification is inadequate. Although he considers the possibility that the parties share a thin concept, he does not take this possibility seriously enough.

Let us turn to Cohen's argument. Though Cohen allows for a thin technical concept of justification, he argues that we have no ordinary concept of it, or that if we do, it is too peripheral to warrant philosophical analysis. The success of his argument rests partly on the truth of the empirical claim that "justified belief" is not regularly used in ordinary discourse. For according to Cohen, the fact that "justified belief" is rarely, if ever, used shows that the concept it expresses is not central to the way we think about ourselves. We can object to this argument on two different grounds. First, we might think that Cohen is wrong about the frequency with which we use "justified belief"—we may actually employ this phrase more often than he claims. But, even if Cohen is correct about this, we still praise others (and ourselves) for having "reasonable," "reliable," and "responsible" beliefs and blame them for having "unreasonable," "unreliable," or "irresponsible" ones. And in so doing we may well be employing a thin concept of justification.

Consider, for example, the ordinary expressions used for acts of free will. Our students may describe these acts as "chosen," "controlled," "voluntary," "uncoerced," "involving decisions," and "involving responsibility."[33] They do not use these terms to express a maximally thick concept of free will, but clearly, there are circumstances in which we would count them as using these terms to express a thin concept of free will. Similarly, suppose that ordinary folks use "reliable," "responsible," "based on good evidence," "believing in a good way," and their complements, instead of "justified belief" and "unjustified belief." Why would we count them as expressing a concept of justification? Because, in using these phrases, they name features of beliefs which we, in our efforts to analyze ordinary concepts, have characterized as relevant to making them justified or unjustified. So, although they do not use "justified belief," they invoke what we recognize to be (proposed) conditions for applying the concept of justified belief. Why would we count them as using "reliable belief," etc., to express a *thin* concept of justification? By using different phrases, i.e., "reliable belief," "responsible belief," etc., they have thickened the thin concept in incompatible ways. But, as long as they also describe the beliefs to which they have applied these terms as, say, "examples of believing in a good way" or "beliefs that have something to do with attaining truths," they will

be employing a thin concept of justification. So, if our students can sometimes use ordinary phrases other than "free will" to express a thin concept of free will, then ordinary folks can sometimes use phrases other than "justified belief" to express a thin concept of justification. Cohen has failed to address this. Nevertheless, I think Cohen and Alston are correct to be suspicious of the internalist-externalist debate. This debate may indeed have been misguided. But, contra Cohen and Alston, this does not warrant utterly abandoning the analysis of justification for the pursuit of other epistemological projects.

Suppose the parties to the debate are treating "justification" as a technical term. We cannot feasibly describe them as having reached complete agreement about the theoretical role of "justification." But, if things are as they appear, and they have agreed that "justification" is to apply to most of our empirical beliefs and all instances of knowledge while having something to do with attaining truths and avoiding falsehoods, they share a thin concept. However, as Cohen has shown, there are equally good but incompatible ways of thickening this thin concept. Internalists will fill in the sketch in one way, externalists in another, but neither is any more correct or less arbitrary. Since there is no single right way to thicken this thin concept, it is indeed misguided to argue about which way of thickening it is the right way. It is wrong-headed to think that there is a *real* notion of justification that one of these precisifications is better at capturing.

Suppose, instead, that they are analyzing a concept that is expressed by a term or terms used in ordinary discourse. Whichever terms we use to express this concept —"justified belief," "reliable belief," "responsible belief," etc.— it will be thin. For though ordinary folks seem to agree that this concept has something to do with reliability, responsibility, and basing one's beliefs on good evidence, there will be no definite answer as to which combinations of these conditions are necessary for its application, or which (short of the whole) are sufficient. Since internalist and externalist precisifications will again be equally correct and equally arbitrary, our ordinary concept of justification will also be too thin to support a meaningful debate between them.

The debate between internalists and externalists has been misguided, but it does not follow that we must abandon the analysis of justification and move on to other projects. To circumvent the debate, we need only precisify our thin concept of justification. That is, parties to the debate need only recognize that they are thickening the thin concept of justification in different ways.[34]

V. The Thin Concept of Intellectual Virtue

The same misgivings that led Alston and Cohen to conclude that we should abandon the analysis of justification have provided part of the impetus for the recent turn to virtue epistemology. The defining feature of virtue epistemology is its focus on the intellectual virtues and vices of agents instead of justification, knowledge, or any other evaluation of belief. Virtue epistemologies focus on the virtues

in the following sense. They define justified belief and/or knowledge in terms of the intellectual virtues and may even eliminate them altogether. They may go on to define the intellectual virtues in terms of states of affairs that are valuable from an epistemic point of view, but they may not define them in terms of justified belief or knowledge.[35] To illustrate, Zagzebski defines justified belief to be "what a person who is motivated by intellectual virtue, and who has the understanding of his cognitive situation a virtuous person would have, might believe in like circumstances."[36] She argues that the intellectual virtues are deep and enduring acquired excellences of a person that involve a motivation to attain truth (and/or understanding) and reliable success in bringing about the end of that motivation. Virtue epistemologies should be contrasted with belief-based theories, which, if they do not eliminate the virtues, define them in terms of justified belief or knowledge. These theories may go on to define justified belief and knowledge in terms of states of affairs that are valuable from an epistemic point of view, but they may not define them in terms of the virtues. So, a belief-based epistemologist might define an intellectual virtue to be a disposition that enables the agent to attain knowledge and define knowledge in terms of, say, tracking the truth. Of course, defining justified belief and knowledge in terms of the intellectual virtues may well change the face of justified belief and knowledge. For if the virtues have motivational components, if they are similar in structure to Aristotelian moral virtues, perceptual knowledge may cease to be paradigmatic.

It is thought that by defining justification in terms of the concept of intellectual virtue — a concept that is sufficiently clear — we can avoid the problems that arose for justification. In other words, if we start with the concept of intellectual virtue and show that justification is merely derivative, we can bypass the fruitless debate. Linda Zagzebski makes exactly this point. She agrees that the internalist-externalist debate is misguided and suggests that virtue epistemology is well-suited to circumvent it because "the blend of internal and external aspects is something that comes with the concept of virtue."[37] She maintains that the concept of virtue has almost always contained both internally accessible elements and internally inaccessible elements. Both are explicitly represented in her own account of intellectual virtue. To illustrate, the virtue of open-mindedness involves both the motivation to be open-minded, which is spawned by the motivation for truth and understanding, *and* reliable success in forming beliefs that are open-minded.[38] The motivation for truth generates the motivation to be open-minded because being open-minded *is* truth conducive and the agent believes this to be the case. So, on Zagzebski's view, intellectual virtues are in part dispositions to have a motive to attain true beliefs and in part reliable mechanisms for attaining true beliefs. Neither component is singly sufficient for the presence of an intellectual virtue.

We have already seen that we need not turn to virtue epistemology in order to circumvent the debate over justification. But, does turning to virtue epistemology provide us with *another* way of avoiding it? That depends on whether the concept of intellectual virtue is sufficiently unproblematic. Zagzebski has maintained that it involves both a motivational and a reliability component. But, other contemporary virtue epistemologists have constructed radically different accounts of intellectual

virtue. According to Alvin Goldman, intellectual virtues are processes (operating with specified parameter values) that produce a high ratio of true beliefs.[39] Memory, sight, and hearing appear on his epistemic evaluator's list of virtues, wishful thinking and guesswork on his list of vices. For Ernest Sosa, one has an intellectual virtue V relative to environment E if and only if "one has an inner nature I in virtue of which one would mostly attain the truth and avoid error in a certain field of propositions F, when in certain conditions C [in environment E]."[40] According to Sosa, fields of propositions and conditions must be the sort that are likely to be repeated in the lives of normal members of one's epistemic community. So, to illustrate, Bob has the virtue of being able to determine the shapes of medium-sized objects relative to the surface of the earth if and only if he has, say, "good eyes and a good nervous system," in virtue of which he would mostly attain true beliefs about the shapes of medium-sized objects, when he sees them in good lighting, at arm's length, and without obstructions here on earth.[41] And, for James Montmarquet, the virtue of open-mindedness is characterized by a desire for truth, but it need not be truth-conducive. On his view, open-mindedness is a tendency to resist dismissing another's ideas on the grounds that they are unfamiliar. He argues that though this resistance is widely believed to be reliable, reliability is not part of what it is to be open-minded.[42] The virtue of open-mindedness need not produce more true beliefs than false ones. Goldman and Sosa clearly employ externalist accounts of virtue, while Montmarquet employs an internalist account. And this is only the beginning of their disagreement about intellectual virtue. Sosa and Goldman maintain that intellectual virtues can be either natural or acquired, while Zagzebski and Montmarquet argue that they must be acquired habits.[43] Sosa implies that the virtues are, or are importantly like, skills, but Zagzebski and Montmarquet deny this.[44] And Zagzebski and Sosa provide different accounts of what makes the virtues valuable — Sosa thinks they are instrumentally valuable; Zagzebski argues that they can be either intrinsically or instrumentally valuable. In short, virtue epistemology appears to be on the brink of its own series of debates. Echoing the advice of Alston and Cohen, should we abandon our analysis of intellectual virtue and move on to some more productive epistemological project?

Certainly not. Granted, these debates would be just as fruitless in virtue epistemology as the internalist-externalist debate was in belief-based epistemology. But, thankfully, thin concepts can inoculate us against this impending conceptual malady. For, like the concept of justification, the concept of intellectual virtue is thin. Zagzebski has precisified it in one way, Sosa and Goldman in another, and Montmarquet in a third. To see that our concept of virtue is thin, let's examine a clear case of intellectual virtue.

Trudy, one of the police force's star detectives, is intellectually thorough with regard to gathering and evaluating evidence. She is disposed to care about gathering the right amount and sort of evidence, and about evaluating it with the appropriate tools and doing so in the appropriate ways. She cares about being thorough because she wants to arrive at true beliefs about "who done it"—she does not want to miss any clues, or be too hasty in analyzing them. When given the opportunity, she exercises the appropriate care and more often than not attains true beliefs about the iden-

tity of the perpetrator. Not only is her thoroughness reliable, but it is an acquired habit that admits of a mean. One could be too thorough by obsessively rechecking evidence, or not thorough enough by glossing over it. Trudy has learned how to hit the mean by imitating her colleagues and teachers and by developing related skills that help her gather and analyze evidence (e.g., the skill of finding drops of blood and hair at a crime scene).[45] Trudy's thoroughness is both instrumentally and intrinsically valuable. It allows her to attain true belief, and it is part of living a good intellectual life. For Sosa, this constitutes a derived virtue; for Goldman, something akin to a method.[46]

When we examine clear cases of intellectual virtue like this one, we find that they exhibit certain features, each of which seems to have something to do with making them intellectual virtues. These features include the following:

(1') an excellence of the agent;
(2') a stable disposition having something to do with truth;
(3') a character trait of the agent;
(4') a disposition to reliably attain true belief and avoid false belief;
(5') the motivation to attain true belief and avoid false belief;
(6') an acquired habit;
(7') a disposition similar to an intellectual skill;
(8') a disposition to hit the mean;
(9') a disposition that is intrinsically valuable;
(10') a disposition that is instrumentally valuable;
(11') a disposition that is not shared by the majority of people in the agent's epistemic community; and
(12') a stable disposition that is somehow related to justified belief and/or knowledge.

When all of these conditions are satisfied, as in Trudy's case, virtue epistemologists will agree that "intellectual virtue" clearly applies; and when none are satisfied, we will agree that it clearly does not apply. We even appear to concur that (1') and (2') are necessary for its application. Still, there will be cases in which some of the conditions are satisfied and others are not, about which virtue epistemologists will disagree. Suppose Bob's disposition clearly satisfies (1')–(3'), (5'), (6'), and (8'), but clearly does not satisfy (4')—it is not truth-conducive. Goldman, Sosa, and Zagzebski would claim that Bob lacks intellectual virtue, while Montmarquet would likely claim that he has it. Likewise, if Bob's disposition clearly satisfied (1'), (2'), (4'), (7'), (10'), and (11'), but clearly failed to satisfy (5'), Montmarquet and Zagzebski would deny that Bob has intellectual virtue, while Goldman and Sosa would claim that he has it. There will be similar disputes over several of the remaining conditions including (6'). These disagreements do not result from a lack of information—again, we all agree on the facts. Rather, as was the case with "religion" and "justification," they result from the meaning of the term in question. The meaning of "intellectual virtue" does not determine which combination of conditions (3')–(11') is necessary for its application, or which, short of the whole, is sufficient. (Sosa seems to think

that the combination of (1'), (2'), (4'), (7'), (10'), and (11') is sufficient, but for Zagzebski the combination of (1')–(6'), (8'), and at least one of (9') or (10') appears to be sufficient.)

Like our concepts of religion and justification, our concept of intellectual virtue admits of borderline cases, like that of Bob, and has fuzzy boundaries. Zagzebski's precisification, which includes (4'), (5'), and (6') as necessary conditions; Sosa's precisification, which includes (4'), but denies that (5') and (6') are necessary; and Montmarquet's, which includes (5') and (6'), but denies that (4') is necessary, are all equally correct and equally arbitrary. Consequently, our concept of intellectual virtue is too thin to make any of these projected disagreements meaningful.

On behalf of Zagzebski, one might respond that even though our technical concept of intellectual virtue is thin, our ordinary concept is fairly thick. That is, our ordinary concept of intellectual virtue has both internally accessible and internally inaccessible elements, and epistemologists who fail to account for this are using "virtue" in a way that deviates from the norm.[47] Let me begin by saying that our ordinary concept of *moral* virtue may be as Zagzebski describes. But, even if it is— even if our concept of moral virtue traditionally contains both of these elements—we still have reason to think that our ordinary concept of *intellectual* virtue is not this thick. I take it that whatever ordinary concept of intellectual virtue we have has, over the last two millennia, been largely, though implicitly, influenced by the Greeks. And Aristotle himself divides the intellectual virtues into two sorts: the contemplative and the calculative.[48] Even if calculative virtues like practical wisdom involve desire, an internally accessible element, contemplative virtues like scientific knowledge and philosophical wisdom do not. According to Sarah Broadie, the virtue of scientific knowledge is an ability to "explain general facts in terms of their causes."[49] It is a capacity to deduce a general fact—e.g., a fact about the heavenly bodies (*Nichomachean Ethics*, 1141b1)—from first principles by means of a syllogism.[50] It is not unreasonable to identify this virtue with the natural faculty of deductive explanation, which allows us to derive so-called "universal facts" from "universal premises." And, indeed, deduction is one of the intellectual virtues named by Sosa and Goldman. If Aristotle's distinction has shaped our ordinary concept of intellectual virtue, then we will be just as likely to name the reliable faculties of deduction and memory "intellectual virtues" as we will be to name practical wisdom and open-mindedness (under their Zagzebskian descriptions). Indeed, I suspect that empirical investigation will confirm that our ordinary use of "intellectual virtue" and its synonyms— "ways of thinking well"—has a rather motley extension. But, even if we identify such diverse dispositions as memory, thoroughness, and, say, mathematical genius as intellectual virtues, it is still reasonable to think that we are employing a single thin concept. For it is still reasonable to think that we will agree that all of these dispositions are excellences of the agent that have something to do with truth.

If, as I have argued, our ordinary and technical concepts of intellectual virtue are thin, then they will not be able to support meaningful debates about whether the virtues have a motivational component, whether they are natural or acquired, or whether they are skills. Fortunately, virtue epistemologists can quash these fruitless debates before they take hold of their discourse by acknowledging that they are

thickening the thin concept of virtue in different ways. In other words, we can avoid these debates by precisifying our concepts of intellectual virtue and/or our concepts of justification and knowledge. Accordingly, Sosa and Goldman can announce that they are interested in externalistic intellectual virtue; Montmarquet, that he is interested in a brand of internalistic virtue; and Zagzebski, that she is interested in a combination of the two. Alternatively, they can precisify the concepts of knowledge and justification that they are defining in terms of the virtues. Indeed, such precisification has already begun. Sosa has argued that true beliefs that result from the exercise of sufficient intellectual virtue constitute *animal* knowledge. For *reflective* knowledge, one's belief must be animal knowledge, but it must also cohere with one's epistemic perspective (one's set of higher-level beliefs about one's lower-level beliefs).[51] Sosa's concept of animal knowledge is externalistic, while his concept of reflective knowledge is arguably a brand of perspectival internalism. Goldman has distinguished between strong and weak justification and has identified beliefs obtained through the exercise of intellectual virtues with the former.[52] On his view, a belief is strongly justified if it is produced by a reliable process; weakly justified if it is produced by an unreliable process, when the agent neither believes this to be so nor has any reliable way of determining it to be so.[53]

Montmarquet has drawn a similar distinction between objective and subjective justification, but has identified beliefs that result from the exercise of intellectual virtues with the latter. On his view, objective justification is externalistic and involves being produced by a reliable process, whereas subjective justification involves the notions of blamelessness and epistemic responsibility.[54] Zagzebski has also identified her concept of justification as deontic because she believes this concept to be closest to that of a right act.[55] And, most recently, John Greco has defined internalistic subjective justification in terms of thinking conscientiously.[56] Once we have identified which thick concepts of intellectual virtue and/or justification and knowledge we are employing, we can substantively debate the nature of the intellectual virtues with those who share our thick concepts. So, Sosa and Goldman can further explore the properties that virtues must have if their exercise is to yield externalistic knowledge or justification. And, Montmarquet, Zagzebski, and Greco can debate the properties virtues must have if they are to be connected to internalistic justification. In short, once we recognize that our shared concept of intellectual virtue is thin, we see that we need not abandon the conceptual analysis of intellectual virtue.

To conclude, I have argued that thin concepts rescue us from fruitless debates about justification and intellectual virtue. Not only can we continue to analyze the concepts of epistemic justification and intellectual virtue, but our analyses will be considerably enlightened once we have acknowledged that we are simply thickening our thin concepts in different ways.[57]

Notes

1. William P. Alston, "Epistemic Desiderata," *Philosophy and Phenomenological Research* 53, 3 (1993): 527–51; Stewart Cohen, "Is There an Issue about Justified Belief?" *Philosophical Topics* 23, 1 (spring 1995): 113–27.

2. James A. Montmarquet, *Epistemic Virtue and Doxastic Responsibility* (Lanham, Md.: Rowman & Littlefield, 1993), ch. 2.

3. Linda Trinkaus Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (New York: Cambridge University Press, 1996), ch. 4.1; Alvin I. Goldman, "Epistemic Folkways and Scientific Epistemology," in *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, Mass.: MIT Press, 1992), 155–75; Ernest Sosa, *Knowledge in Perspective: Selected Essays in Epistemology* (New York: Cambridge University Press, 1991), 284–88.

4. I am indebted to Wayne Riggs for his comments on an earlier version of the introduction.

5. Alston, "Epistemic Desiderata," 541.

6. Ibid., 528–30.

7. The basic aims of inquiry include attaining truth and avoiding falsehood, being reflectively aware of one's epistemic situation, and having an integrated system of belief.

8. Alston, "Epistemic Desiderata," 533; my emphasis.

9. Ibid., 533.

10. Ibid., 536.

11. Ibid., 536.

12. Alston acknowledges that many of the issues discussed under the rubric of justification will survive the transition to desiderata-based epistemology.

13. Cohen, "Is There an Issue about Justified Belief?" 121–22.

14. Ibid., 116.

15. Ibid.

16. Ibid., 120.

17. Ibid.

18. William P. Alston, *Philosophy of Language* (Englewood Cliffs, N.J.: Prentice Hall, 1964). The same example appears in William P. Alston, "Vagueness," in Paul Edwards ed., *Encyclopedia of Philosophy*, vol. 8 (New York: Macmillan, 1967), 218–21.

19. Alston, "Vagueness," 220.

20. Ibid., 219.

21. Ibid.; Alston, *Philosophy of Language*, 88.

22. Alston, *Philosophy of Language*, 89; my emphasis.

23. Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (New York: Macmillan, 1953), sections 66–71. Also see Alston, "Vagueness," 220; Michael P. Lynch, *Truth in Context: An Essay on Pluralism and Objectivity* (Cambridge, Mass.: MIT Press, 1998), 63.

24. Peter Geach and Max Black, eds., *Translations from the Philosophical Writings of Gottlob Frege*, 3rd. ed. (Oxford: Basil Blackwell, 1985), 139.

25. Timothy Williamson, *Vagueness* (New York: Routledge, 1994), 41.

26. Geach and Black, *Translations,* 139.

27. Lynch, *Truth in Context,* 72.

28. Alston, *Philosophy of Language*.

29. Rosanna Keefe and Peter Smith, eds., *Vagueness: A Reader* (Cambridge, Mass.: MIT Press, 1996), 5.

30. David Lewis, *On the Plurality of Worlds* (New York: Basil Blackwell, 1986), 244fn32. Also see Keefe and Smith, *Vagueness,* 23–35; Williamson, *Vagueness,* 142–64.

31. Roughly, a vague sentence is supertrue if it is true under all permissible precisifications (e.g., Orthodox Judaism is a religion); (super)false if it is false under all permissible precisifications (e.g., gardening is a religion), and neither true nor false if it is true under some, but false under others (e.g., Communism is a religion).

32. Richard Foley may disagree. See Richard Foley, *The Theory of Epistemic Rationality* (Cambridge, Mass.: Harvard University Press, 1987).

33. Thanks to Amy Allen for this example.

34. Thanks to Matthias Steup for this point.

35. According to Sosa, the virtues are instrumentally valuable because they allow us to attain true beliefs. See *Knowledge in Perspective*, 225. Zagzebski distinguishes between two different sorts of virtue theories: In a happiness-based theory, the virtues are instrumentally valuable; in a motivation-based theory, they are intrinsically valuable.

36. Zagzebski, *Virtues of the Mind,* 241. She thinks knowledge is a state of belief arising out of acts of intellectual virtue. See 270–71.

37. Ibid., 332.

38. Ibid., 185–86.

39. Goldman, "Epistemic Folkways and Scientific Epistemology."

40. Sosa, "Intellectual Virtue in Perspective," in *Knowledge in Perspective*, 284.

41. Sosa, "Reliabilism and Intellectual Virtue" in *Knowledge in Perspective*, 139.

42. Montmarquet, *Epistemic Virtue*, 25.

43. Sosa claims that there are both fundamental and derived virtues; Goldman claims that there are both belief-forming processes and methods. See Sosa, "Intellectual Virtue in Perspective," 278; Goldman, "Strong and Weak Justification," in *Liaisons*, 128–30; Zagzebski, *Virtues of the Mind*, 102–106; Montmarquet, *Epistemic Virtue*, 26.

44. Sosa, "Virtue Perspectivism: A Response to Foley and Fumerton," *Philosophical Issues 5: Truth and Rationality* (1994), 30–33; Zagzebski, *Virtues of the Mind*, 106–116; Montmarquet, *Epistemic Virtue*, 34.

45. To be intellectually thorough, one must also have acquired the motivation to be thorough and the ability to size up situations properly. On the acquisition of the moral virtues, see Nancy Sherman, *The Fabric of Character*: *Aristotle's Theory of Virtue* (New York: Clarendon Press, 1989).

46. If Sosa and Goldman maintain that this is not a paradigm case of intellectual virtue, then virtue epistemology is in even worse shape—we are not even operating under the rubric of a shared thin concept.

47. Thanks to Abrol Fairweather for this point.

48. Aristotle, *Nicomachean Ethics*, trans. David Ross (New York: Oxford University Press, 1992), VI.1.

49. Sarah Broadie, *Ethics With Aristotle* (New York: Oxford University Press, 1991), 223.

50. Ibid. Also see *Nicomachean Ethics*, 139b27–35, 1141b31–33.

51. Sosa, "Intellectual Virtue in Perspective," 287–290.

52. Goldman, "Epistemic Folkways and Scientific Epistemology," 157 and fn 7.

53. Goldman, "Strong and Weak Justification."

54. Montmarquet, *Epistemic Virtue*, ch. 6.

55. Zagzebski, *Virtues of the Mind*, 232–41.

56. John Greco, "Agent Reliabilism," in James E. Tomberlin, ed., *Philosophical Perspectives 13: Epistemology* (Atascadero, Calif.: Ridgeview, 1999), 289.

57. I am deeply grateful to Linda Zagzebski, William Alston, and Wayne Riggs for their comments on earlier drafts of this essay. I would also like to thank Matthias Steup, Abrol Fairweather, Andrew Ward, Samuel Levey, Christine Thomas, Jon Ellis, Amy Allen, Walter Sinnot-Armstrong, John Hawthorne, Seth Shabo, and my colleagues at California State University Fullerton.

8

VIRTUES AND RULES IN
EPISTEMOLOGY

John Greco

Virtue theories in ethics are often presented as an alternative to deon-
tological theories such as Kant's. One important issue here concerns
the existence of moral rules and their relation to moral action. Specifically, it is com-
monly claimed that it is an advantage of virtue theories that they do not commit us
to moral rules. The reason this is an advantage, it is argued, is that it is unclear
whether such rules play any important role in our moral lives.[1] In this essay I want
to argue in a similar way with respect to theories of epistemic evaluation. That is, I
want to argue that virtue theories in epistemology hold an advantage over deonto-
logical theories in epistemology, and precisely because the former need not under-
stand epistemic justification in terms of epistemic rules or norms.

To clarify my thesis, I need to say something about what I mean by "epistemic
justification" and something about what I mean by "deontological theories." First,
by "epistemic justification" I mean the kind of justification that is involved in knowl-
edge. William Alston has argued that there are a variety of "epistemic desiderata."[2]
In Nicholas Wolterstorff's terminology, there are a variety of "truth-relevant mer-
its."[3] It is no part of my thesis that deontological theories do not capture one or more
of these various merits. Again, my thesis is about the kind of justification involved in
knowledge. Another way to get at the target concept is to say that it is the kind of
justification that turns true belief into knowledge, Gettier problems aside, perhaps.
Thus it is what Roderick Chisholm called "positive epistemic status," Alvin Plan-
tinga calls "warrant," and Ernest Sosa calls "aptness."[4]

Second, by deontological theories I have in mind theories of justification that are
rule-based or norm-based. We may think of cognitive rules or norms as conditional
statements whose antecedents specify cognitive states of the believer, and whose con-
sequents specify further cognitive states as permitted, required, or prohibited. More
exactly, their antecedents specify features of cognitive states, since by nature rules are
general. They tell us to behave in certain ways given conditions of certain kinds.
John Pollock has argued that cognitive norms are permission rules, having the form
"In conditions X, it is permissible to do Y."[5] Alternatively, one might hold that they

are prescription rules of the form "In *X* do *Y*," or proscription rules of the form "In *X* do not do *Y*." Or perhaps cognitive rules come in various forms, including all of those just mentioned. Putting such controversies aside, the main idea of deontological theories is that justified belief is a function of correct cognitive rules or norms. More exactly,

> DJ. *S* is epistemically justified in believing *p* if and only if *S*'s believing *p* is licensed by correct cognitive rules.

Theories of this kind are modeled on deontological theories in ethics. Thus, for example, Kant held that an action is right just in case it is licensed by a correct "maxim," or rule of action. In ethics and in epistemology, deontological theories can differ with regard to which rules are correct, why they are correct, how one can know that they are correct, whether one must know that they are correct, and along other dimensions as well. But the main idea of deontological theories is that some relevant merit, moral or epistemic, is a function of whether one's activity is licensed by some relevant set of rules.

Putting all this together, my thesis is that rule-based theories of epistemic justification are false. Most of the essay will be devoted to establishing that this is so. A consequence of this thesis is that we must understand the kind of justification involved in knowledge in some other way. At the end of the essay I will make some brief suggestions in favor of a virtue approach, arguing that virtue theories of justification are able to avoid the kinds of objection that I bring against deontological theories. In fact, I will argue, virtue theories in epistemology become especially attractive in light of the kinds of considerations that are here raised against deontological theories.[6]

I. Weak and Strong Deontological Justification

The argument against DJ must proceed in two parts. This is because there are two major kinds of deontological theory and they must be argued against in different ways. We may understand the division by noting a distinction that is important in Kant's moral theory. Specifically, we can distinguish between (a) action that is merely in accordance with duty and (b) action that is for the sake of duty. Kant gives an example to illustrate. First, a grocer might decide to give a child correct change because this is what self-interest requires. In other words, it is good business to deal with all of one's customers honestly. Alternatively, the grocer might give the child correct change because this is what duty requires. In the second case, the rule or maxim behind the grocer's action is a law of morality rather than a council of prudence, and so Kant says that the action is "for the sake of duty," or alternatively, "out of reverence for the moral law." According to Kant, actions that are in accordance with duty are "right" in the sense that they violate no moral law, but only actions that are for the sake of duty have moral worth.

Thus in ethics as in epistemology we may recognize a variety of desiderata or merits. Discussing a different example, Kant writes:

> To help others where one can is a duty, and besides this there are many spirits of so sympathetic a temper that, without any further motive of vanity or self-interest, they find an inner pleasure in spreading happiness around them and can take delight in the contentment of others as their own work. Yet I maintain that in such a case an action of this kind, however right and however amiable it may be, has still no genuinely moral worth . . for its maxim lacks moral content, namely, the performance of such actions, not from inclination, but *from duty*.[7]

We may therefore use Kant's distinction to specify two kinds of deontological merit. Specifically, weak deontological merit requires that one's action not violate any relevant rule. Strong deontological merit requires that one's action be *governed* by the right rules. Roughly, the stronger condition requires the causal efficacy of the rules in question. One's action must be the result of *following* the rules, as opposed to merely being describable by them.

Applying these considerations to epistemic justification, we may distinguish two versions of DJ above.

DJ(W). A belief is epistemically justified in the weak sense if and only if it violates no correct cognitive rule.

DJ(S). A belief is epistemically justified in the strong sense if and only if it is governed by correct cognitive rules; i.e., one's belief is the result of following such rules.

Having made this distinction, we are now is a position to argue against both weak and strong deontological theories. Here are the two arguments, briefly stated.

The main argument against weak deontological theories is as follows: Where knowledge is concerned, etiology matters. For example, in cases where one's knowledge depends on evidence, it is required that one believes what one does *because* of the evidence in question. However, etiology does not matter for weak deontological justification. One has this sort of justification so long as one does not in fact violate any relevant rule; as with Kant's weak sense of right action, it does not matter *why* one does not violate the rule. But then weak deontological justification is not epistemic justification, or the kind of justification required for knowledge.

The main argument against strong deontological theories is this: Strong deontological justification requires that one's beliefs be governed by correct cognitive rules. But it is possible to have knowledge even if one's cognition is not rule governed at all. Therefore, knowledge does not require strong deontological justification. Here is the argument a bit more formally, where K is the proposition that one has knowledge and R is the proposition that relevant parts of one's cognition are rule governed.

(1) DJ(S) > nec (*K* > *R*)
(2) poss (*K* & not-*R*)

Therefore,

(3) not-DJ(S)

Next I will present two arguments in favor of the possibility thesis stated in premise (2). The first is due to Thomas Reid and concerns the possibility of immediate knowledge of the world. According to Reid, God could have created human beings with different cognitive faculties than the ones we actually have. For example, God could have created us with an immediate and infallible knowledge of the material world, with no need for sensory input to ground our beliefs. Such knowledge would not be governed by cognitive rules, because it would not work by going from cognitive inputs to cognitive outputs and therefore would not require cognitive rules to govern such transitions. But it would be knowledge nonetheless. In fact, Reid thought that parts of human perception actually work this way, and recent empirical studies tend to support him on this point.

The second argument for premise (2) invokes connectionist models of cognitive processing. According to this argument, it is an empirical question, concerning a contingent matter of fact, whether various parts of human cognition are governed by cognitive rules. But however such empirical questions are answered, it will still be the case that paradigm instances of human cognition count as knowledge. For example, if cognitive science tells us that connectionist models of perception are true, and that perception is in fact not rule governed, we will not conclude that there is no perceptual knowledge. Therefore, it is possible that knowledge is not rule governed.

These are the main arguments that I will raise against weak and strong deontological theories. The rest of the essay is details. In Section 2, I will develop the argument against weak deontological theories of justification, or theories that endorse DJ(W). In Section 3, I will develop the argument against strong deontological theories of justification, or theories that endorse DJ(S). In Section 4, I will briefly argue that virtue theories of epistemic justification can avoid the objections raised against both kinds of deontological theory. Again, a virtue approach in epistemology becomes especially attractive in light of the considerations here raised against deontological theories.

2. The Argument against Weak Deontological Theories

The argument against weak deontological theories claims that etiology is important for knowledge. In other words, in cases of knowledge it matters why one believes what one does. One way to see this is to note the distinction between (a) *having* good reasons for what one believes, and (b) believing *for* good reasons. For example, anyone who knows the axioms of arithmetic has good reasons for believing a theorem in the system. But unless one puts two and two together, so to speak, one does not believe the theorem in question on the basis of one's good reasons. We can imagine, for example, that someone believes a true theorem for no reason at all, or even that

one believes a true theorem for bad reasons. In either case, the person in question will have good reasons so long as he believes the relevant axioms, but the person will not believe for good reasons.[8]

The next point is that knowledge requires believing for good reasons, and not just having good reasons. We may consider two examples to illustrate this. First, consider the math student who knows all the relevant axioms but doesn't see how the axioms support a theorem that must be proven on the exam. Eventually he reasons fallaciously to the theorem and believes it on the basis of his fallacious reasoning. Surely he does not know that the theorem is true.

Second, suppose that Charlie is a wishful thinker and believes that he is about to arrive at his destination on time. He has good reasons for believing this, including his memory of train schedules, maps, the correct time at departure and at various stops, etc. However, none of these things is behind his belief — he does not believe what he does *because* he has these reasons. Rather, it is his wishful thinking that causes his belief. Accordingly, he would believe that he is about to arrive on time even if he were not. Again, it is clear that Charlie does not have knowledge that he will arrive on time. In an important sense, it is merely an accident that he has good reasons for his belief, since his belief is not affected by those reasons.

In both of the cases described, the person in question has good reasons for his or her belief but does not believe for those reasons. Remember, these are not supposed to be cases of over-determination of belief. Rather, we are describing persons whose good reasons play no part at all in their believing what they do. But then clearly the people in question do not have knowledge. As we said earlier, in cases of knowledge it matters *why* one believes what one does.

And now for the final point in the argument against DJ(W): Weak deontological justification requires only that one have good reasons for what one believes, and not that one believe for good reasons. Thus in each of the cases above, the person in question violates no correct cognitive rule in believing as they do. Remember, those rules specify that some belief is permitted, or necessary, or prohibited given antecedent cognitive conditions. But since, by hypothesis, the persons in question have good reasons for what they believe, their beliefs will not violate any such rule. The people in the examples are like Kant's self-interested grocer: they "do the right thing," but not for the right reasons. But then weak deontological justification is too weak for epistemic justification, or the kind of justification required for knowledge. The kind of justification required for knowledge requires believing for good reasons, whereas weak deontological justification does not.

At this point someone might raise an objection. Correct cognitive rules, it might be said, require that one believes that one's beliefs are based on good evidence, or that one's evidence supports one's beliefs, or that one's evidence makes one's beliefs probable, or the like. Since the people in the examples do not have such beliefs about their evidence, it is false that they do not violate any relevant cognitive rule. One problem with this objection is that we can change the examples so that the people in question do have the relevant beliefs about their evidence. So long as these beliefs about their evidence are not involved in the production of their further beliefs, the people in the revised examples will fail to have knowledge for just the same reason

as in the original examples. Specifically, although they have good reasons for what they believe, they do not believe for good reasons.[9]

We may conclude that weak deontological theories are false. Strong deontological theories of epistemic justification, however, can accommodate the considerations raised previously concerning the etiology of knowledge. This is because strong deontological justification requires that one's belief be governed by correct cognitive rules. Put another way, it requires that one *follow* the rules rather than merely not violate them. This condition on justification insures that one believes *for* good reasons in cases of justified belief, since those reasons are involved in the antecedents of the relevant cognitive rules. In cases where one's belief is deontologically justified in the strong sense, one follows the rules of good reasoning and thereby bases one's belief on good reasons.

Strong deontological theories, therefore, offer a more plausible account of epistemic justification than weak deontological theories.[10] To show that strong deontological justification is nevertheless not epistemic justification, a different argument is needed.

3. The Argument against Strong Deontological Theories

The argument in Section 2 concluded that weak deontological justification is too weak for epistemic justification. The argument in this section will be that strong deontological justification is too strong. The short form of the argument was put as follows: Strong deontological justification requires that one's beliefs be governed by correct cognitive rules. But it is possible to have knowledge even if one's cognition is not rule governed at all. Therefore, knowledge does not require strong deontological justification.

The key to the argument is clearly premise (2), or the claim regarding the possibility of knowledge that is not rule governed. The remainder of this section will be devoted to the task of establishing that possibility. What is important to emphasize, however, is that it is only the possibility that needs to be established for the argument to go through. Accordingly, I will not be arguing for the empirical thesis that relevant parts of human cognition are in fact not rule governed. Rather, I will be taking advantage of the fact that strong deontological theories are philosophical theses, and therefore make a claim that carries necessity. Such theories do not claim that knowledge just happens to be governed by correct cognitive rules. On the contrary, they claim that knowledge must be governed by correct cognitive rules, and that this is a matter of conceptual or perhaps metaphysical necessity. But then if it is possible to have knowledge that is not rule governed at all, the claim about necessity is false.

As noted above, I will present two arguments for the possibility thesis in premise (2): one from Reid concerning the possibility of immediate knowledge, and one from connectionist theory concerning the possibility of non-rule-governed cognitive processing. Before considering these arguments in detail, however, it will be necessary to say more about the nature of cognitive rules and also more about what it

means for cognition to be governed by such rules. The danger in arguing against strong deontological theories is that we build too much into the notion of "being governed by a rule," thereby scoring only a hollow victory by showing that knowledge need not be rule governed in that sense. My strategy for avoiding this danger will be to consider several senses of "being governed by a rule," and to argue that it is possible that knowledge is not rule governed even in the weakest of these senses. I turn to these preliminaries now.

a. The Nature of Cognitive Rules

We said that cognitive rules are conditional statements whose antecedents specify features of cognitive states and whose consequents specify further cognitive states as permitted, required, or prohibited. This characterization raises a question regarding which features of cognitive states figure into cognitive rules. A natural answer is that they are those features that we have beliefs about, for how else could we follow the rules except by believing that their antecedents are fulfilled? It turns out that this answer won't do, however, because it entails an infinite regress of beliefs: If every belief requires a further belief about antecedents being fulfilled, then there will be no end to the beliefs that are required.[11] What is needed, then, is a way to follow cognitive rules without first having to have beliefs about their antecedents being fulfilled. Pollock suggests the following solution.

> Now that we understand how epistemic norms work in guiding our reasoning, it is easy to see that they must be internalist norms. . . . In general, the circumstance-types to which our norms appeal in telling us to do something in circumstances of those types must be directly accessible to our automatic processing systems. The sense in which they must be directly accessible is that our automatic processing system must be able to access them without our first having to make a *judgment* about whether we are in circumstances of that type.[12]

> It seems clear that directly accessible properties must be in some sense "psychological," but I doubt that we can say much more than that from the comfort of our armchairs.[13]

Pollock's idea is that we can be aware that some features of cognition obtain even if this awareness in not by means of beliefs or judgments about those features. What he has in mind are various aspects of our beliefs and experience that he calls "internal" and "psychological," but he does not go on to say precisely what these labels mean. One way to interpret "internal" in the present context is "intrinsically conscious." The idea would be that, by their very nature, various aspects of our cognitive states are at the level of conscious awareness. That we are being appeared to redly, or that a belief has a particular narrow content, would be candidates for such status. On this suggestion we are necessarily conscious of certain aspects of our cognitive states, and it is in that sense that we are aware of them even when we do not make judgments about them.

This suggestion is too strong for present purposes, however, because not all cog-

nition takes place at the level of consciousness. And if some stretch of cognitive activity is not conscious at all, then it cannot be governed by means of features of it that are conscious. Perhaps we should say that the features of cognition that figure into cognitive rules are "potentially conscious" and be satisfied to leave "potentially" unanalyzed for present purposes. What the rule theorist means to pick out, clearly enough, are the various aspects of our cognitive states that figure into sound reasoning, good judgment, and the like. And these would seem to include various syntactic, semantic, and qualitative properties that are at least often at the level of conscious awareness.

The current suggestion is full of problems and controversy. Perhaps most important, it is not clear what properties of beliefs and experience count as potentially conscious in the relevant sense. But at this point in the argument we should be as generous as possible. Remember, we are trying to characterize strong deontological theories so that they at least get off the ground, and this requires that we have a characterization of cognitive rules that makes being governed by such rules possible. And this, we have seen, requires that we can be aware that the antecedents of such rules are fulfilled in some way other than by making judgments about them. Hence we have Pollock's notion of directly accessible properties, which I have glossed in terms of potentially conscious properties.

Pollock argues that cognitive rules must be stated exclusively in terms of directly accessible properties, but that might be stronger than what is required for present purposes. For it seems that we can solve the aforementioned regress problem so long as *some* cognitive rules can be followed without first having to have beliefs that their antecedents are fulfilled. Accordingly, we can expand the notion of "potentially conscious property" as follows. Some features of things are potentially conscious by their very nature: for example, various aspects of our beliefs and experiences. Other features, however, are potentially conscious in the sense that we can be "conscious of" them by having representations of them. If we expand the notion of "potentially conscious property" in this way, then we can understand cognitive rules as operating on features of things that are represented by our cognitive states as well as on features of our cognitive states themselves. For example, it seems perfectly reasonable to count the following as a cognitive rule: "If a reliable authority is telling you that p, then it is permissible to believe that p." In other words, there is no need to restrict ourselves to, "If you believe that a reliable authority is telling you that p. . ." Let us adopt this broader notion of "potentially conscious property" and understand cognitive rules as operating on features of cognitive states that are potentially at the level of consciousness, and also on features of things that our cognitive states are about. Remember, we are trying to make the notion of "being governed by a cognitive rule" as broad as possible, and so we should make our notion of a cognitive rule as broad as possible.

Even on this broader characterization, however, someone might object that we are making the notion of a cognitive rule too strong. Why couldn't properties that are clearly not potentially conscious, even in the broader sense, go into the antecedents of cognitive rules, so long as our cognition is somehow responsive to them? For example, why can't it be a cognitive rule that, given some pattern of light

on the retina (even where we do not have a belief about this), one ought to believe that it is raining? The reply is that we are looking for a sense of "cognitive rule" that is relevant to strong deontological theories of justification, and this requires a kind of rule that we can in some sense respect, or take into account, or follow. And this, in turn, requires that we be in some sense *cognitively* aware that the antecedents of such rules are fulfilled, even if we have expanded our notion of "awareness of *X*" far beyond the notion of "having beliefs about *X*." As we will see next, we can articulate weaker and stronger notions of "being governed by a rule." But not just any notion will do. For example, we will want to distinguish "being governed by a rule" from "being subject to a causal law." Consider that any sort of physical thing is subject to causal laws, and in that sense is "responsive" to the fact that the laws' antecedents are fulfilled. But not any sort of thing follows rules, which implies at least some kind of cognitive awareness. In sum, we want a notion of a cognitive rule that can govern our cognition, and we do not want to stretch the latter notion beyond any sense relevant for deontological justification.

b. Senses of "Being Governed by a Rule"

This brings us to our next task, which is to articulate some different senses of "being governed by a rule." As was said earlier, my strategy will be to determine several senses of "being governed by a rule" that are compatible with strong deontological justification, and to argue that it is possible that knowledge is not rule governed even in the weakest of these senses.

We may turn again to Kant's moral theory for the strongest possible sense of "being governed by a rule." As we saw above, Kant requires for moral worth that one's action be done for the sake of duty, and this requires that one have as the maxim of one's action some moral law. In several places Kant seems to require that the maxim of one's action be explicitly represented. Thus we have:

> Now an action done from duty has to set aside altogether the influence of inclination, and along with inclination every object of the will; so there is nothing left able to determine the will except objectively the *law* and subjectively *pure reverence* for this practical law, and therefore the maxim of obeying this law even to the detriment of all my inclinations.[14]

Presumably one cannot have reverence for a law unless one represents it to oneself, and in fact this is Kant's view.

> What I recognize immediately as law for me, I recognize with reverence, which means merely consciousness of the *subordination* of my will to a law. . . . Immediate determination of the will by the law and consciousness of this determination is called '*reverence*'. . . .[15]

The picture we get here is that of a person explicitly considering some rule of action and acting out of reverence for that rule. Whether this is in fact Kant's considered

view, it provides a model for the strongest sense of "being governed by a rule"; that is, conscious consideration of an explicitly represented rule, which then becomes the motive behind one's action.

We may determine weaker senses of "being governed by a rule" by eliminating various elements in this characterization. To do so it will be helpful to make some distinctions. The first is between occurrent and dispositional representations. The second is between conscious and nonconscious representations. To say that one has a representation occurrently is to say that one has it at the moment. To say that one has a representation dispositionally is to say that one has a disposition to have it occurrently in relevant circumstances. A different distinction is between conscious and nonconscious representations. To have a representation consciously is to be presently aware of some representational content. To have a representation nonconsciously is to somehow have the representation presently, but in a way that one is not presently aware of its content. It is controversial whether there are any nonconscious representations in this sense, but clearly some people think that there are. For example, Freudians claim that some of our beliefs are repressed and yet continue to have effects on our actions. Now they could not mean that we have such beliefs dispositionally, so that the conscious/nonconscious distinction is just the occurrent/dispositional distinction. This is because it is part of the Freudian doctrine that we are *not* disposed to have repressed beliefs consciously. In fact, the doctrine is that we are disposed not to have them consciously. Whether or not nonconscious representations actually exist, it is clear that our two distinctions cut across each other. Thus we might have occurrent conscious representations, occurrent nonconscious representations, dispositional conscious representations, and dispositional nonconscious representations.

With these distinctions in mind we can more clearly specify the strongest sense of "being governed by a rule" attributed to Kant above. Thus we have:

RG1. *S*'s action *A* is governed by rule *R* if and only if *S* has an occurrent, conscious representation of *R*, and this representation is causally involved in *S*'s doing *A*.

We may specify weaker senses by eliminating various aspects of this characterization. Thus we have:

RG2. *S*'s action *A* is governed by rule *R* if and only if *S* has an occurrent, nonconscious representation of *R*, and this representation is causally involved in *S*'s doing *A*.

RG3. *S*'s action *A* is governed by rule *R* if and only if *S* has a dispositional representation of *R*, and this disposition is causally involved in *S*'s doing *A*.

An even weaker sense of "being governed by a rule" can be specified by means of a further distinction. We can follow Terence Horgan and John Tienson by distinguishing between (a) having a disposition to represent a rule, and (b) having a dis-

position to act in a way that the rule specifies, where the latter notion does not involve either occurrent or dispositional representation of the rule.

> Morphological possession of intentional content M is a matter of the cognitive system's being disposed, by virtue of its persisting structure rather than by virtue of any occurrent states that are tokens of M, to undergo state transitions that are systematically appropriate to content M—and to do so, at least much of the time, without generating a token of M during the process. Morphological content differs from occurrent representational content (e.g., occurrent belief) because it involves the cognitive system's persisting structure, rather than occurrent tokening of M. Morphological content differs from dispositional representational content (e.g., dispositional belief) . . . because the relevant dispositions associated with morphological content involve tendencies other than the tendency to generate token representations with that content.[16]

Applying the idea of morphological possession to the notion of a rule, we have:

> RG4. S's action A is governed by rule R if and only if S has a disposition to act in the way specified by R, and this disposition is causally involved in S's doing A.

One qualification at this point is necessary, as it will become important later. Elsewhere Horgan and Tienson argue that a subject follows a cognitive rule only if her transition from input cognitive states to output cognitive states goes through all of the various steps that the rule specifies.[17] The idea here is clear enough: If S gets from point A to point B in some way other than the steps that rule R specifies, then S did not get from A to B by following R. RG4 should be read with this understanding in mind.

This weakest sense of "being governed by a rule" seems to be the one intended by Pollock when he describes how our beliefs are guided by epistemic norms.[18]

> We know how to reason. That means that under various circumstances we know what to do in reasoning. This can be described equivalently by saying that we know what we should do. Our epistemic norms are just the norms that describe this procedural knowledge. . . . They describe an internalized pattern of behavior that we automatically follow in reasoning, in the same way we automatically follow a pattern in bicycle riding.[19]

As in RG4, Pollock's sense of "being governed by a rule" does not require that the rule be represented, even dispositionally.

> It is this process that I am calling "'being guided by the norm without having to think about the norm.'" This may be a slightly misleading way of talking, because it suggests that somewhere in our heads there is a mental representation of the norm and that mental representation is doing the guiding. Perhaps it would be less misleading to say that our behavior is being guided by our procedural knowledge and the way in which it is being guided is described by the norm.[20]

Whereas the strongest sense of "being governed by a rule" required conscious, oc-current representation of the rule, our weakest sense requires no representation of the rule at all.

Is RG4 thereby rendered too weak for present purposes? In other words, is it too weak for a theory of strong deontological merit, with the latter's emphases on *following* a rule, or *respecting* a rule, or acting *for the sake of* a rule? The following considerations suggest that it is too weak.

First, consider that various animal behaviors might count as following a rule in the sense of RG4. For example, it has been observed that birds feeding from various sources will eventually conform to an algorithm for maximizing the intake of food.[21] But it seems a stretch to say that the birds are "following a rule" in any sense that is relevant to deontological merit. Since the birds are not even potentially aware of the rule they are "following," it is only in an extended sense, seemingly irrelevant to any kind of intellectual credit, that their behavior is governed by the rule. A similar point applies to human behavior. Scientists trying to model human cognition have found that simple perceptual tasks such as pattern recognition require algorithms of astounding complexity—far more complexity than the average perceiver could handle in his or her explicit calculations.[22] Again, it is therefore a stretch to say that perceivers are "following" the algorithms in any but the most extended sense. A more apt description of both the birds and the humans is that their cognitive activity is governed by causal laws, and that this results in cognitive dispositions that are at most *describable* by an algorithm; they do not result from anything like *using* the algorithm, or respecting it, or following it.

I believe that this kind of consideration tells against the appropriateness of RG4. In effect, it shows that RG4 is beginning to stretch the notion of "being governed by a rule" beyond any sense relevant for deontological merit. I will resist the temptation to strengthen RG4, however, because there are reasons for thinking that knowledge need not be rule governed even in this very weak sense. If these reasons are sound, then we may conclude that knowledge need not be rule governed in any allowable sense.

c. Reid's Argument for the Possibility of Knowledge That Is Not Governed by Rules

The purpose of the present section is to establish the possibility of knowledge that is not governed by cognitive rules. The first argument for this possibility comes from Thomas Reid. According to Reid, it is a matter of fact that human beings perceive physical objects by means of sensory experience. Given the way human cognition actually works, it is necessary that there be some physical interaction with the object perceived, which interaction gives rise to sensations, which in turn give rise to perceptual beliefs about the object.

> Although there is no reasoning in perception, yet there are certain means and in-struments, which, by the appointment of nature, must intervene between the object and our perception of it; and, by these, our perceptions are limited and regu-

lated. First, if the object is not in contact with the organ of sense, there must be some medium which passes between them. Thus, in vision, the rays of light; in hearing, the vibrations of elastic air; in smelling, the effluvia of the body smelled — must pass from the object to the organ; otherwise we have no perception. . . . The [physical] impression made upon the organ, nerves, and brain is followed by a sensation. And, last of all, this sensation is followed by the perception of the object. . . .[23]

According to this description, human perception is rule governed in the sense of RG4. Human perception involves dispositions to go from certain kinds of sensations to certain kinds of beliefs about physical objects, and therefore human perception is governed by the rules describing these dispositions. However, Reid argues, it is merely a contingent fact about human cognition that our perception is governed by such rules. God could have made us so that we perceive physical objects immediately, without need of the sensations that in fact precede our perceptions.

We might, perhaps, have been made of such a constitution as to have our present perceptions connected with other sensations. We might, perhaps, have had the perceptions of external objects, without either impressions upon the organs of sense, or sensations. Or, lastly, the perceptions we have might have been immediately connected with the impressions upon our organs, without any intervention of sensations.[24]

The last possibility described by Reid is the one that is interesting for our purposes. Reid's idea is essentially as follows: As a matter of fact, human perception takes place by means of two separate dispositions. The first is a disposition to go from physical "impressions" on the sense organs to mental "sensations" caused by those impressions. The second is to go from these sensations to beliefs about physical objects. What Reid is suggesting is that God could have eliminated the "middleman" in the process. There is no reason why physical impressions on our sense organs could not have led immediately to beliefs about the objects that cause those impressions. If this were the case, then we would have law-governed perception but not rule-governed perception in the sense of RG4.

Perception would be law governed, on Reid's view, just as any natural activity is governed by the laws of nature. But since these laws would govern the formation of beliefs on the basis of physical impressions on the sense organs, they would not count as cognitive rules in the sense defined in Section 3a and employed in RG4. It will be remembered that cognitive rules must specify "potentially conscious properties" in their antecedents. But physical impressions on the sense organs do not count as potentially conscious properties, even on the very broad sense that was allowed in Section 3a. As defined there, such properties were understood to include (a) various features of our cognitive states that are at least sometimes at the level of conscious awareness, and (b) any properties that are represented by our cognitive states. But physical impressions, such as those that light makes on the retina or sound makes on the eardrum, fall into neither of these categories in the typical case. So even on our weakest notion of "being governed by a rule," and employing our broadest notion of

a "potentially conscious property," the perception that Reid imagines would not be governed by cognitive rules.

And now we need only make explicit what is already implicit in Reid's discussion, namely, that such perception would nevertheless give rise to knowledge. Suppose we come across some class of God's creatures (human or otherwise) whose perception works in the way that Reid describes. Suppose also that the perception of such creatures is as extensive and as reliable as ours is, or even more so. Such creatures are able to navigate their environment precisely and reliably, are able to accurately identify wide ranges of objects as having wide ranges of properties, and so on. Add also that such creatures reason appropriately from their perceptions, give due consideration to counter evidence, and so on. Clearly we would not judge that they do not have knowledge of their physical environment. Rather, we would say that they acquire knowledge differently than we do. We may conclude, therefore, that knowledge need not be rule governed in the sense of RG4.

The creatures we have been discussing are not imaginary on Reid's view. Rather, Reid thought that certain ranges of human perception work in just the way described.

> When I see an object, the appearance of which the colour of it makes, may be called *the sensation*, which suggests to me some external thing as its cause; but it suggests likewise the individual direction and position of this cause with regard to the eye. I know it is precisely in such a direction, and in no other. At the same time, I am not conscious of anything that can be called *sensation*, but the sensation of colour. The position of the coloured thing is no sensation; but it is by the laws of my constitution presented to the mind along with the colour, without any additional sensation.[25]

Reid goes on to describe how this might happen.

> We have reason to believe, that the rays of light make some impression upon the *retina*; but we are not conscious of this impression . . . . Now, this material impression, made upon a particular point of the *retina*, by the laws of our constitution, suggests two things to the mind—namely, the colour and the position of some external object . . . . And since there is no necessary connection between these two things suggested by this material impression, it might, if it had so pleased our Creator, have suggested one of them without the other. Let us suppose therefore, since it plainly appears to be possible, that our eyes had been so framed as to suggest to us the position of the object, without suggesting colour, or any other quality: What is the consequence of this supposition? It is evidently this, that the person endued with such an eye, would perceive the visible figure of bodies, without having any sensation or impression made upon his mind.[26]

Finally, Reid draws his conclusion:

> If we suppose, last of all, that the eye hath the power restored of perceiving colour, I apprehend that it will be allowed, that now it perceives figure in the

very same manner as before, with this difference only, that colour is always joined with it.

In answer, therefore, to the question proposed, there seems to be no sensation that is appropriate to visible figure, or whose office it is to suggest it.[27]

Reid thinks that normal human beings perceive the position and visible figure of objects without the benefit of sensations. I will not argue that Reid is correct on this point of fact, nor is it necessary for me to do so. Rather, I am concerned to establish a possibility, and it does seem to me that Reid has established that much. In other words, he has established that perception could work in the way that he describes, whether or not it does work that way in humans. This is because perception is essentially discrimination: To perceive physical objects as having certain properties is to discriminate the object and the properties from other possibilities. We already assume that human perception is underwritten by physical processes that allow for such discrimination. Whether it is by means of light on the retina, sound on the eardrum, or "the effluvia of the body smelled," we assume that minute variations in physical events affect our sense organs differently, thereby allowing the various discriminations we make by means of the five senses. As Reid points out, eliminating an intermediate stage in the process would not affect the ability to make those discriminations, so long as the relationships between physical inputs and perceptual outputs remained unaffected. Now suppose that further investigation confirmed that Reid is right—that we do visually perceive position and figure, but not by means of visual sensations. In that case it would be established that such perception is not rule governed in the sense of RG4. But clearly we would not give up the idea that we can know the positions and figures of physical objects just by looking.

It is interesting to note that Reid is not alone in postulating this sort of cognition. On the contrary, in many ways Reid's discussion anticipates current investigations into blindsight and other forms of nonconscious perception. Blindsight is now a widely documented phenomenon, occurring in subjects with damage to the primary visual cortex. Subjects who are blind in some part of their visual field can nevertheless discriminate size, shape, location, and/or orientation of objects in the blind part of the field despite the absence of any conscious visual experience. In typical cases subjects insist that they cannot see the stimuli and therefore cannot answer questions about it. However, in "forced-choice testing" these subjects perform at levels that are better than chance and sometimes remarkably high. Speculation about how blindsight works is also reminiscent of Reid's discussion. Since damage in blindsighted subjects is to the visual cortex rather than to the eyes themselves, it is hypothesized that information from the eye still reaches the brain, although bypassing the mechanisms normally responsible for conscious visual experience.[28]

More recently it has been argued that nonconscious perception is a part of normal cognitive functioning.[29] This is in fact a controversial issue, but again, whether nonconscious perception actually exists is not relevant for present purposes. What is relevant is the possibility, and that seems to be established well enough. Given the way that standard perception actually works, it seems a small stretch that information about the environment could be carried to the brain without sensory awareness—

all we need imagine is that the path from physical stimulus to conscious belief does not go through sensory experience. Finally, consider that both the phenomenon of blindsight and Reid's perception of visible figure are radical versions of nonconscious perception, in that they hold that no sensory awareness at all is involved in the relevant processes. But perception would fail to be rule governed in the sense of RG4 so long as it is *partially* a function of non-cognitive inputs. Suppose that perception is a function of (a) initial sensory inputs, (b) existing beliefs in the system with relevant content, and (c) initial non-cognitive inputs such as Reid's "physical impressions." If this is the case, then there will be perceptual dispositions that are not specifiable in purely cognitive terms, and therefore perception that is not rule governed in the sense of RG4.

Whether nonconscious perception actually exists seems a matter of contingent fact, to be decided by empirical investigation into the actual workings of human cognition. Now suppose that such perception does exist. Suppose also that people reason appropriately from such perceptions, give due consideration to counter evidence, and do all the other things that we think are required for positive epistemic status. Why should we deny epistemic justification to such beliefs, simply for lack of preceding conscious experience? In fact, we would not deny justification. On the contrary, we would say that perceptual knowledge arises differently than we had supposed.[30]

So far we have been arguing for the possibility of knowledge that is not rule governed by considering the possibility of cognitive processing with non-cognitive inputs. The basic idea has been that the cognitive processes that lead to belief might be processes that go from non-cognitive physical states to cognitive states, rather than from cognitive states to cognitive states, as rule governance in the sense of RG4 would require. Another way that knowledge might not be rule governed in the sense of RG4 is that cognitive processing always goes from cognitive inputs to cognitive outputs, but not by means of dispositions that are describable by cognitive rules. Put another way, the transitions from cognitive inputs to cognitive outputs might not be the result of dispositions that are describable on the cognitive level. Connectionist models of cognitive processing demonstrate this possibility.

d. The Second Argument for the Possibility of Non–Rule-Governed Knowledge

It is now commonplace to think that connectionist models of cognitive processing imply that cognition is not governed by rules. I will begin this section by giving a brief description of such models. Second, I will review some reasons for thinking that the models imply that cognition is not governed by rules, paying special attention to the sense of "governed by rules" stipulated in RG4. The argument for the possibility thesis in premise (2) will then be as follows: It is an empirical question, concerning a contingent matter of fact, whether connectionist models of human cognition are true. But however this question is decided, we will continue to judge that paradigm cases of cognition, for example, simple perceptual judgments, count as cases of knowledge. Therefore, it is possible that knowledge is not governed by rules.

Connectionist systems consist of numbers of simple but connected units that can be "activated" or "excited" to some degree. The units are set up so that the activation of each affects the activation of others to which it is immediately connected. In typical systems the interaction among units affects connection "strengths" or "weights" over time, so that the effect of one unit on another is increased or inhibited on the basis of prior interaction between the two units. Processing takes place in a connectionist system when an initial pattern of activation is supplied to the input units. This original activation sets off activity among the various units of the system until stability in the system is achieved. The pattern of activation over the output units then represents the system's "answer" to a proposed problem. Thus in connectionist systems the input and output units, or in some cases patterns of activation over these, are given interpretations for the purposes of the task being modeled. In typical cases a system will contain "hidden" units as well. Hidden units need not be interpreted themselves. Rather, their purpose is to effect appropriate transitions from interpreted inputs to interpreted outputs.

We may now see why connectionist models have been thought to imply that cognition is not rule governed. First, it is clear that on such models cognitive processing need not take place by means of *explicitly represented* rules: Although some connectionist models do in fact employ explicitly represented rules, the more interesting models do not. Rather, connectionist processing takes place by means of the interactions among various connected units in the system, and these interactions are governed only by physical laws. Initial activation, together with initial connection strengths, cause a pattern of activity that eventually settles the system into the state of highest entropy.[31] Clearly, therefore, on connectionist models cognitive processing need not be rule governed in the sense of RG1 or RG2, each of which require an occurrent representation of the rules in question. Neither must processing be rule governed in the sense of RG3, since there need be no disposition to represent relevant rules either. If cognition must be rule governed at all on connectionist models, then it will have to be in the weak sense of RG4. I now turn to two arguments from Horgan and Tienson, which establish that not even this is the case.

First, recall what it means for cognition to be rule governed in the sense of RG4. In general, S's action A is governed by rule R if and only if S has a disposition to act in the way specified by R, and this disposition is causally involved in S's doing A. In the specific case of cognition, action A refers to a cognitive action and rule R refers to a cognitive rule, where cognitive rules are understood as above. Specifically, cognitive rules specify features of cognitive states in their antecedents, and then specify further cognitive states as permissible, required, or prohibited. Putting all this together, S's cognition is rule governed in the sense of RG4 just in case S's cognition is the result of dispositions to go from earlier cognitive states to later cognitive states in a way specified by some set of cognitive rules.

The next thing to note is that on some connectionist models cognition *is* rule governed in the sense of RG4. The important point, however, is that this is not so on all connectionist models. The reason for this is that connectionist processing takes place on the subrepresentational level. In other words, it takes place by means of the myriad of interactions between individual, directly connected units in the system. But

the dispositions produced by subrepresentational processing need not mirror any disposition that is describable at the cognitive level.

Horgan and Tienson make the point as follows.

> According to [the present connectionist model], the human cognitive system's transitions from one total cognitive state to another leap fairly large cognitive gaps in single steps that are not decomposable, via classicist "boxology," . . . into computationally subservable cognitive "baby steps" (either serial ones, or simultaneous ones, or both). Although there exist algorithms for computing the system's cognitive-transition function (because, *ex hypothesi*, the cognitive transitions are tractably computable), the system itself does not execute any such representation-level algorithm; i.e., the various physical processes in the system do not realize all the distinct symbol-manipulation steps whose collective physical realization would constitute execution of a given algorithm. Even though the system might execute algorithms at one or more *subrepresentational* levels of description (as does a standard connectionist network), its processing does not conform to PRL rules—i.e., programmable rules that refer solely to those formal/mathematical features of representations that play representational roles.[32]

Horgan and Tienson's discussion is aimed at classical AI's assumptions regarding cognitive processing; that is, that such processing must take place solely by means of operations on the syntactic/formal properties of representations, and in a way that can be implemented in a physical system. However, their argument can be carried over straightforwardly for our purposes. The essential point is that processing that takes place on the non-cognitive level of nodes and connections need not be describable by rules stated at the cognitive level. But if this is the case, then the cognitive processing in a connectionist network need not give rise to dispositions that are describable by such rules, and so rule governed in the sense of RG4.

The following example illustrates the present point.

> Consider standard decision theory, viewed as a putative psychological model of human deliberation and choice, rather than a normative theory. According to this model, a deliberating agent will choose an action with maximum expected utility. The expected utility of an envisioned act is a certain kind of weighted sum: the sum of respective numerical values the agent assigns to the various envisioned potential outcomes of that act, with each value weighted by the agent's subjective probability of the given outcome's resulting from that act.
>
> Now, one way such a decision-making system might work would be computational. For instance, it might actually calculate this weighted sum, for each envisioned act, then compare the totals for the acts and calculate which act or acts have maximal expected utility, and then pick an act with maximal expected utility. But there is another possibility, without representation-level computations: There are various beliefs and desires at work in the system, with various strengths. They all enter the hopper at once and interact directly—somewhat in the manner of a complex combination of interacting physical forces in a planetary system, with the various bodies exerting mutual gravitational influence on one another. The way they interact is via a kind of "resolution of forces," where

the forces get resolved in such a way that the cognitive system eventually settles on an alternative with maximal expected utility.[33]

Again, the cognitive activity in the example would be *law* governed, although it would not be *rule* governed in the sense of RG4.

It is also important to realize that connectionist networks could, in principle, sub-serve non-algorithmic cognitive transitions, even though each node in the net-work updates its own activation, locally, in accordance with an algorithm. The local updatings of individual nodes need not be parallel baby step in some algo-rithm *over representations*—some set of programmable rules for manipulating and transforming complex representations on the basis of their representation-level structure.[34]

This first argument from Horgan and Tienson depends on the idea that the non-cognitive rules involved in connectionist processing need not mirror cognitive rules in all their "baby steps." A second argument they make against rule-governed cog-nition does not depend on this point. In the passage below, Horgan and Tienson are discussing their preferred model of connectionist processing. According to that model, human cognition is not "tractably computable," meaning that it cannot be physically implemented by means of a set of programmable, representation-level rules. In the passage "CTF" stands for "cognitive transition function" and "TCS" stands for "total cognitive state."

What does it mean for a human CTF not to be tractably computable? The CTF itself can be construed as an enormous set of ordered pairs, each of which asso-ciates a single TCS with a set of one or more successor TCSs. One way to spec-ify this function would be via a huge (possibly infinite) list, with each ordered pair in the CTF specified by a separate entry. Such a list, even if finite, would be truly gargantuan—far too big to itself constitute a set of programmable rules. Thus, a tractably computable CTF, because of the enormous number of distinct cognitive transitions it includes, would have to be fully specifiable in some way other than via a brute list.
   What classicism assumes, of course, is that a human CTF is specifiable via some set of *general laws* over cognitive states; each determinate cognitive transi-tion type is supposedly just a particular instance of one of these laws, and the CTF delineated by the laws is supposedly tractably computable. Thus, if the CTF is not completely specifiable by such general laws, it will not be tractably computable.[35]

Horgan and Tienson's point is that the CTF need not be so specifiable if it is imple-mented by a connectionist network. Precisely because cognitive transitions in a con-nectionist network are effected by means of *sub*representational laws, there is no guarantee that those transitions will also be describable by representation-level laws. Of course there will be a cognitive transition *function*; that is, a function that is specifiable at the level of representations. But this function need not be specifiable in any general way at the representational level. This point is important in the con-

text of Horgan and Tienson's discussion because they are concerned to reject classical AI's assumption that cognitive transitions can be physically implemented by means of representation-level rules. If there is no way to specify the cognitive transition function of a system in a general way, then that will not be the case. But their point is important for our purposes as well. Putting the point in terms of cognitive states rather than in terms of representations: If the dispositions of the system need not be describable in a general way at the level of cognitive states, then that amounts to saying that its dispositions need not be describable by cognitive *rules*. And that means that the activity of the system need not be governed by cognitive rules in the sense of RG4, since "governed by cognitive rules" in that sense just means "describable by cognitive rules."

Horgan and Tienson give several arguments for concluding that their model gives an accurate account of human cognition. But as I have said above, it is not my purpose here to defend any such empirical thesis. Rather, the relevant point for my purposes is that Horgan and Tienson have described a possible model of human cognition. Now suppose that their possible model is correct and that cognitive science establishes this beyond any doubt. Suppose that paradigm cases of perception, inductive reasoning, memory, and the like arise in accordance with the model and that we all accept that this is the case. For example, we all believe that walking into the room and seeing that there are chairs there is the result of just the kind of connectionist processing that the model invokes. Surely we would not judge that therefore no one knows that there are chairs in the room. We would not judge, that is, that we have no perceptual knowledge, no knowledge by induction, and so on. On the contrary, we would judge that we have learned some interesting details about how such knowledge in fact arises. Clearly, then, the concept of knowledge does not require that knowledge be rule governed. Strong deontological theories of epistemic justification are false.[36]

4. A Virtue Approach to Epistemic Justification

In this last section I want to briefly argue that virtue theories of epistemic justification provide an attractive alternative to deontological theories. We have been thinking of deontological theories in epistemology as modeled after Kantian rule-based theories in ethics. By contrast, virtue theories in epistemology are modeled after virtue theories in ethics. This main idea is best understood in terms of a thesis about the direction of analysis. Just as virtue theories in ethics try to understand the normative properties of actions in terms of the normative properties of moral agents, virtue theories in epistemology try to understand the normative properties of beliefs in terms of the normative properties of cognitive agents. Hence, virtue theories in ethics have been described as person-based rather than act-based, and virtue theories in epistemology have been described as person-based rather than belief-based.[37]

What is the nature of the intellectual virtues? As in moral theory, this is a controversial question on which various positions are possible. Thus Kant thought that

moral virtue was primarily a function of proper motivation. In a famous passage, Kant writes, "A good will is not good because of what it effects or accomplishes—because of its fitness for attaining some proposed end: it is good through its willing alone—that is, good in itself." And again,

> Skill and diligence in work have a market price; wit, lively imagination, and humour have a fancy price; but fidelity to promises and kindness based on principle (not on instinct) have an intrinsic worth. In default of these, nature and art alike contain nothing to put in their place; for their worth consists, not in the effects which result from them, not in the advantage or profit they produce, but in the attitudes of mind—that is, in the maxims of the will—which are ready in this way to manifest themselves in action even if they are not favoured by success.[38]

According to Aristotle, however, the moral virtues involve both a motivational component and a reliable success component: a virtuous person is motivated toward the good, but is also successful in achieving it. Hence Aristotle agrees that moral virtue is "concerned with choice," but also writes,

> every virtue or excellence both brings into good condition the thing of which it is the excellence and makes the work of that thing be done well. . . . Therefore, if this is true in every case, the virtue of man also will be the state of character which makes a man good and which makes him do his own work well.[39]

An Aristotelian account of the moral virtues, therefore, combines considerations of both responsibility and reliability. Like Kant's good will, the morally virtuous person is praiseworthy in that her character displays an appropriate motivational orientation. But on Aristotle's account, the morally virtuous person is also reliable, in that she reliably achieves the moral goods that she is motivated to pursue.

Now it seems to me that the Aristotelian model is the better one for theories of epistemic justification, since it seems to me that knowledge requires both responsibility in one's cognitive conduct and reliability in achieving epistemic ends. But however we decide this issue, the main point is that virtue theories define epistemic justification in terms of the normative properties of persons, i.e., the stable dispositions or character traits that constitute their intellectual virtues, however these are to be understood. Here is an example of how such a definition might be constructed. First, we can define subjective praiseworthiness and objective reliability as follows:

V(SJ). S is subjectively praiseworthy in believing p if and only if S's believing p results from the intellectual dispositions that S manifests when S is motivated to believe the truth.

V(OJ). S is objectively reliable in believing p if and only if S's believing p results from dispositions of S's intellectual character that reliably produce true belief.

We may then define epistemic justification as follows:

V(EJ). A belief *p* is epistemically justified for *S* if and only if *S* is both subjec-
tively praiseworthy and objectively reliable in believing *p*.

And now we are in a position to make the main point of this section: Whether we
understand epistemic justification in terms of motivational orientation, reliable suc-
cess, or both, a virtue approach is neutral regarding whether justified belief is rule
governed. More specifically, virtue theories are neutral regarding whether the cog-
nitive dispositions that give rise to epistemic justification are grounded in rule-
governed behavior.

Here we may follow Sosa by distinguishing between a disposition and the basis
for that disposition.

> Of course one and the same virtue might have several different alternative possi-
> ble grounds or bases. Thus the disposition to roll down an incline if free at its top
> with a certain orientation, in a certain environment (gravity, etc.), may be
> grounded in the sphericity and rigidity of an object, or alternatively it may be
> grounded in its cylindricality and rigidity. Either way, the conditional will obtain
> and the object will have the relevant disposition to roll. Similarly, Earthians and
> Martians may both be endowed with sight, in the sense of having the ability to tell
> colors and shapes, etc., though the principles of the operation of Earthian sight
> may differ widely from the principles that apply to Martians, which would or
> might presumably derive from a difference in the inner structure of the two
> species of being.[40]

In a similar manner, we could understand the dispositions involved in proper
motivation and reliable success as being themselves grounded in conformance to
cognitive rules. If the rules are of the right kind, and if they reliably produce be-
lief in *S*'s environment, then *S*'s beliefs will satisfy the definitions above in virtue of
S's cognitive activity being governed by those rules. Thus a virtue approach to epi-
stemic justification is consistent with the hypothesis that knowledge-producing cog-
nition is rule governed. But a virtue approach is also consistent with the alternative
hypothesis that knowledge-producing cognition is not rule governed. The defini-
tions above say nothing about *how S*'s dispositions arise, and so it leaves it open that
S's dispositions might be grounded in cognitive architecture that does not involve
cognitive rules. And this is just what we want if the architecture question is an em-
pirical one.

Finally, we may note that virtue theories also avoid the objection raised above
against weak deontological theories. The problem for weak deontological theories
was that they do not discriminate between having good reasons and believing for
good reasons. More generally, in cases of knowledge etiology matters: It is required
that the knower form her belief in the right way, or at least maintain it in the right
way. Virtue theories easily accommodate this consideration, for they require that
knowledge be grounded in virtuous character. In cases of inferential knowledge,
this will involve dispositions to form one's beliefs on the basis of reliable evidence.
In cases of non-inferential knowledge, this will involve dispositions to form one's
beliefs in other ways that are both responsible and reliable. Hence virtue theories

clearly fall into the class of what Goldman calls "genetic theories."[41] What distinguishes virtue theories from other genetic theories is that, as Aristotle says, "the moving principle is in the agent himself."[42]

In conclusion, I have argued against both weak and strong deontological theories of justification, and I have argued that virtue theories of justification can avoid the objections raised against each. The main objection raised against weak deontological theories is that they fail to take into account the causal etiology of belief. For example, such theories fail to distinguish between having good reasons and believing for good reasons. But whether a belief qualifies as knowledge is partly a matter of etiology, and so weak deontological theories are too weak. The main argument against strong deontological theories is that they say too much about causal etiology. Specifically, such theories require that knowledge be governed by rules, whereas our intuitions about which cases count as knowledge support no such requirement. Therefore, strong deontological theories are too strong. In contrast, virtue theories make causal etiology matter, requiring that in cases of knowledge belief is the result of virtuous cognitive character. However, virtue theories need not require that knowledge be governed by rules. On the contrary, they can make this an empirical question about the basis of intellectual virtue, rather than a philosophical question about the conditions for knowledge.[43]

Notes

1. For example, see G. E. M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33, 124 (1958): 1–19; Alasdair MacIntyre, *After Virtue* (Notre Dame, Ind.: University of Notre Dame Press, 1981); and John McDowell, "Virtue and Reason," *Monist* 62, 3 (1980): 331–50. For a valuable discussion of relevant literature, see Robert Roberts, "Virtues and Rules," *Philosophy and Phenomenological Research* 51, 2 (1991): 325–43.

2. William Alston, "Epistemic Desiderata," *Philosophy and Phenomenological Research* 53, 3 (1993): 527–51.

3. Nicholas Wolterstorff, "Epistemology of Religion," in John Greco and Ernest Sosa, eds., *The Blackwell Guide to Epistemology* (Oxford: Blackwell Publishers, 1999).

4. Roderick Chisholm, *The Theory of Knowledge*, 2nd. ed. (Englewood Cliffs, N.J.: Prentice Hall, 1977); Alvin Plantinga, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993); and Ernest Sosa, *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991).

5. John Pollock, *Contemporary Theories of Knowledge* (Totowa, N.J.: Rowman & Littlefield, 1986).

6. Deontological theories are commonly characterized as duty-based or obligation-based, rather than as rule-based or norm-based. However, I prefer to use the present terminology for a number of reasons. First, rule theorists in both ethics and epistemology tend to speak in terms of obligation and duty, so rule-based theories and duty-based theories in both fields tend to overlap. A case in point is Kant's moral theory, which I here use as a paradigmatic instance of the kind of theory that I am targeting. Second, I am influenced by Anscombe's classic essay, which argues that the close relationship between rule-based theories and duty-based theories is not accidental. Anscombe argues that in moral philosophy, the language of duty and obligation is rooted in a conception of ethics that centers on the notion of moral law. It is only within this theoretical framework, she argues, that such language is really intelligible. See Anscombe, "Modern Moral Philosophy." Nevertheless, it is

important to note that my objections to deontological theories are directed at the notion of cognitive rules of a certain sort and that, therefore, any theory that is committed to such rules is subject to the same objections. I want to thank Terence Horgan and Matthias Steup for pressing these points to me.

7. Immanuel Kant, *Groundwork of the Metaphysic of Morals*, H. J. Patton, trans. (New York: Harper and Row, 1964), 66.

8. For an extended discussion of this distinction and its importance, see Robert Audi, *The Structure of Justification* (Cambridge: Cambridge University Press, 1993), especially chapter 7.

9. Alvin Goldman makes a similar point in "What Is Justified Belief?" in George Pappas, ed., *Justification and Knowledge* (Dordrecht: D. Reidel, 1979). Another problem with the current objection is that it invokes a requirement that would lead to skepticism. See my "Agent Reliabilism," *Philosophical Perspectives, 13, Epistemology,* James Tomberlin, ed. (Atascadero, Calif.: Ridgeview Press, 1999).

10. This would explain why I have defended such a theory in the past. See my "Internalism and Epistemically Responsible Belief," *Synthese* 85 (1990): 245–77 and "Virtues and Vices of Virtue Epistemology," *Canadian Journal of Philosophy* 23 (1993): 413–32.

11. Pollock, *Contemporary Theories*.

12. Ibid., 133.

13. Ibid., 135.

14. Kant, *Groundwork*, 68–9.

15. Ibid., 69n.

16. Terence Horgan and John Tienson, "Connectionism and the Commitments of Folk Psychology," *Philosophical Perspectives, 9, AI, Connectionism and Philosophical Psychology*, James Tomberlin, ed. (Atascadero, Calif.: Ridgeview Press, 1995), 132.

17. "In order for cognitive processing to 'conform to PRL rules,' in the sense intended, it is not enough that the system's input/output function match the one subserved by the rules; rather, the system must convert inputs to outputs by going through all the steps the rules prescribe." *Connectionism and the Philosophy of Psychology* (Cambridge, Mass.: MIT Press, 1996), 176n14.

18. It is also the sense that I intended in "Internalism and Epistemically Responsible Belief" and "Virtues and Vices of Virtue Epistemology," where I largely followed Pollock on this point.

19. Ibid., 131.

20. Ibid.

21. A number of relevant studies are discussed in John E. R. Staddon, "Optimality Theory and Behavior," in John Dupre, ed., *The Latest on the Best: Essays on Evolution and Optimality* (Cambridge, Mass.: MIT Press, 1987).

22. J. Hochberg, *Perception* (Englewood Cliffs, N.J.: Prentice Hall, 1978).

23. Thomas Reid, *Philosophical Works*, H. M. Bracken, ed., 2 vols. (Hildesheim: Georg Olms, 1983), 186a–b.

24. Ibid., 187b.

25. Ibid., 145a.

26. Ibid., 146a.

27. Ibid., 146a–b.

28. See L. Weiskrants, *Blindsight: A Case Study and Implications* (Oxford: Clarendon Press, 1986).

29. For an informative review of relevant literature, see Paul Lewicki, Thomas Hill, and Maria Czyzewska, "Nonconsious Acquisition of Information," *American Psychologist* 47, 6 (1992): 796–801.

30. I have been making the present argument in terms of perceptual knowledge. Simi-

lar arguments could be made, however, regarding the possibility of other kinds of immediate knowledge, for example, memory knowledge and intuitive knowledge of simple necessary truths.

31. William Bechtel, "Connectionism and the Philosophy of Mind: An Overview," in William Lycan, ed., *Mind and Cognition* (Oxford: Basil Blackwell, 1990), 254.

32. *Connectionism and the Philosophy of Psychology*, 55.

33. Ibid., 56.

34. Ibid., 56–7.

35. Ibid., 64.

36. Wayne Riggs raises the objection that I have not considered an important kind of deontological theory: namely, one that defines epistemic justification in terms of rules of conduct rather than in terms of cognitive rules. I believe that there are two decisive considerations against such theories, however. The first is that such theories do not discriminate finely enough between justified and unjustified belief. For example, one might engage in grossly irresponsible conduct and nevertheless be in a position to have knowledge as a result, for example through perception. This sort of consideration is brought out forcefully by William Alston in "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives, 2, Epistemology,* James Tomberlin, ed. (Atascadero, Calif.: Ridgeview Press, 1988). The second problem with such theories is that they put off the objections raised against strong deontological theories only for a moment. Specifically, the question will arise concerning how one knows that the antecedents of the rules of conduct are satisfied. Depending on how one answers, we will either be off on a regress or we will have to acknowledge the role of cognitive rules as described above.

37. Linda Zagzebski, *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996).

38. Kant, *Groundwork*, 62, 102.

39. Aristotle, *The Nicomachean Ethics*, David Ross, trans. (Oxford: Oxford University Press, 1980), Book II, ch. 6.

40. Sosa, *Knowledge in Perspective*, 141.

41. Goldman, "What Is Justified Belief?"

42. Aristotle, *Nicomachean Ethics*, Book III, ch. 1.

43. I would like to thank Robert Audi, Terence Horgan, Hilary Kornblith, Ralf Muller, Wayne Riggs, Ernest Sosa, Margaret Walker, and especially Matthias Steup for their comments on earlier drafts of this essay.

9

MUST KNOWERS BE AGENTS?

Linda Zagzebski

I. Act and Agency

Acts are performed only by certain kinds of beings—agents (from the Latin *agens*: that which is acting), although that tells us neither what an act is nor what an agent is, only that "act" and "agent" are defined correlatively. Ethics is a field primarily concerned with the evaluation of human agents, their acts, and the consequences of their acts, but again, this does not tell us what ethics does, only that certain things it does are connected to certain other things it does. In particular, it does not tell us what the scope of ethics is, which might turn out to be greater or less than we think.

I assume that most human beings are agents at a fairly early age, and it is possible that some other animals are agents also, although I will not discuss animal agency here. An agent is the kind of being that acts. To act is to exert power and, at least typically, to bring about a certain kind of effect through the exercise of that power. I say "at least typically" because successfully bringing about an effect of a certain kind might be treated as constitutive of agency—presumably, a power is not a power unless it is reliably effective. But we are somewhat forgiving about the degree of efficacy required of human agents since human power is obviously not infallibly efficacious, yet no one takes that fact to be incompatible with the existence of human agency. An agent is still an agent if she occasionally is ineffective in bringing about the effect associated with her power, and she is exercising her agency even on such occasions. The extent to which she can be ineffective and remain an agent will be one of the issues addressed in this essay.

An agent, then, is a being that brings about certain kinds of effects through the exercise of a power. But to do so is not sufficient for acting since many artifacts and inanimate substances bring about similar effects, but they do not act. We may even say they have "powers," although it is likely that a chemical *agent* (note the term) has a power only in an extended sense. In the strict sense only conscious beings have powers. It is interesting to consider why that should be the case. What difference does it make if a being with the capacity to produce effects is a conscious agent?

That is, what difference does it make if a causal capacity is a power in the sense I am using rather than a mere capacity? If all that counts is the causal connection itself, it should not matter. Clearly, it does matter to ethics, but it is not clear why it does since the class of all acts includes more than the class of acts subject to moral evaluation. The issue I want to begin investigating here is whether it matters to epistemology.

An important way to classify acts in order to bring out their susceptibility to moral evaluation is to classify them according to degree of conscious awareness and/or control. In such a classification one extreme would consist of fully deliberate acts, acts preceded by conscious deliberation and choice. These acts fall within a larger class of intentional acts that includes some that are not preceded by deliberation and choice, and a still larger class of acts that are non-intentional but are nonetheless subject to moral evaluation. Aristotle called this larger class the voluntary [*hēkōn*]. Non-voluntary acts are acts that are not subject to moral evaluation, at least not in the sense that the agent of such an act is praiseworthy or blameworthy for doing it. But a non-voluntary act is still an act. It is not like the event of a chemical agent producing the effect it produces. There is still a difference between an act, whether voluntary or not, which must be performed by a conscious agent, and the non-acts of non-agents that also bring about effects in the world.[1] The nature of that difference is another matter I want to begin investigating here.

Some philosophers have distinguished agent causation from event causation to mark this difference in the way effects are related to their causes. A chemical reaction is said to be fully explained by the sequence of events leading up to it; the concept of event causation is sufficient to give an account of a chemical event. In contrast, the acts of agents are a kind of event that allegedly cannot be fully explained by relations between events. The concept of the agent is critical in explaining why and how events of this kind occur. If event causation is to be distinguished from agent causation, that is because it is maintained that the cause of an act is not an event, not even the event of an agent's performing the act; it is the agent herself.[2]

The causal relation is notoriously resistant to analysis, but attempts to analyze it often construe it in terms of counterfactual conditionals. This attempt is problematic, but even some of the theorists who would not reduce causation to counterfactuals maintain that causal propositions entail counterfactuals. Counterfactual conditionals may therefore be helpful in illuminating the causal relation. If so, agent causation would have implications concerning what the agent would do in counterfactual circumstances. And since agent causation is connected to the idea of a power, that is probably what we would expect since power no doubt implies the production of the same type of effect in a range of circumstances, including many that are non-actual. But are the relevant counterfactuals entailed by the possession of a power, or are they merely the sign of it? In addressing the issue of what determines the degree of causal efficacy required to be an agent, I will investigate the relevance of counterfactual conditions to agency.

The term "agent causation" was popularized in contemporary discourse by Roderick Chisholm, who attributes the idea to Thomas Reid, although a form of the idea exists much earlier, perhaps even in Aristotle.[3] Reid proposed that agent causation is a more basic notion than causation *simpliciter* since an understanding of cau-

sation presupposes an understanding of power. Our notion of active power is more conceptually basic than that of causation, Reid claimed; it is presupposed by our knowledge of ourselves as rational and moral agents.[4] The idea of causation is derived from the idea of agency and responsibility. Presumably, we would not have acquired the idea of causation if we were merely witnesses to nature. Reid maintained that agency even appears in the operation of our intellectual powers, including those operative during perception. The faculty of perception is an original power of the mind. The true cause of perception is the agent exerting this power to produce an effect. Sensations and impressions are not the real causes, much less the objects perceived.

I mention Reid's strong view on the place of agency in perception not to endorse it, but to highlight a question that I think deserves more attention in epistemology: What is the place of acts and agency in the acquisition of epistemic states, particularly those that are evaluatively positive, such as justified belief, responsible belief, and knowledge? In this essay I am particularly interested in the relevance of agency to the acquisition of knowledge. It is uncontroversial that processes and events of some kind lead up to a human being's coming to know something. Coming to know is an event, and the issue of what causes that event is important, not only because some form of the causal theory of knowledge might be true, but because understanding the cause of something almost always helps us to understand it better. Is knowledge best understood on the model of event causation or on the model of agent causation? That is another question I want to begin investigating in this essay.

While knowledge and justification are often connected with causation in the contemporary literature, agent causation is rarely mentioned. Although Aristotle and Aquinas referred to the "act of knowing," nowadays knowing is more commonly construed as a state rather than as an act.[5] This may be due in part to the fact that perceptual knowledge is commonly taken to be the paradigm and perception is usually understood as a relatively passive state, or at any rate, as a state one acquires prior to the exercise of one's agency. It may also be partly due to the fact that the range of acts and the corresponding range of agency has narrowed significantly in modern philosophy for reasons that derive from ethics rather than the philosophy of perception. The importance of the act in Kant was gained at the price of narrowing its scope to a single tightly circumscribed deliverance of will. The mind itself is now often viewed as a passive information processor rather than as an active agent. This view has led to a shift in the prototype of the act in modern philosophy. We no doubt find it curious that in Aquinas the act par excellence is a mental act since in contemporary discussions the prime example typically given of a basic act is the raising of one's arm. So these days when we think of an act we usually think either of an act of will or of a willed bodily movement. Cognitive and perceptual acts only make sense on this view if preceded by acts of will. The broader Aristotelian category of the voluntary and the even broader category of acts both voluntary and non-voluntary have generally disappeared from discussion. Granted, the fully intentional act is in many ways the most interesting kind of act, but that does not mean that there are no interesting differences between acts and non-acts. If there are such differences, the epistemic arena is one place in which we would expect them to appear.

So far I have identified three sets of questions about agency that I want to begin investigating:

1. What are the conditions for being an effective agent? What determines that an agent is effectively exercising her agency on a particular occasion? Must she be reliable? Is her efficacy determined by what she is able to do in counterfactual circumstances?
2. Is there any important difference between an effect arising from the act of an agent, whether voluntary or non-voluntary, and events brought about by a non-agent? In particular, does it make any significant difference to epistemology?
3. Is knowledge best understood on the model of event causation or on the model of agent causation?

II. Agency and Counterfactual Conditions

What are the conditions for being an effective agent? If I am right in my conjecture that effectiveness is part of the concept of an agent, then the conditions for being an effective agent are the conditions for being an agent. An effective agent is one whose acts are successful in reaching their end.[6] Presumably effectiveness comes in degrees, and so my degree of effectiveness as an agent is partly a matter of the proportion of my successes to failures in achieving the ends of my acts. But my effectiveness as an agent is also a matter of the extent to which my successes can be credited to *me* rather than to something else. An effective agent is one who reaches her end *because of* her act, the exercise of her power. This rules out both accidental success and success that is non-accidental but due to something other than the agent.[7] An effective cook produces a high proportion of good dishes over mediocre ones and does so because of what *she* does in exercising her cooking ability rather than by chance or because someone else is guiding her every step of the way. An effective cook gets the credit for her culinary successes. An effective teacher produces a high proportion of students knowledgeable or skillful in the subject of his course, and he does so because of what he does in exercising his teaching ability rather than by chance, or because his students are simply bright enough to learn on their own or for some other reason. An effective teacher gets the credit for his students' success in learning.

The same point applies to our effectiveness as moral agents. An effectively compassionate agent is one who produces a high proportion of successes at alleviating suffering and who does so because of the exercise of his own power in reducing suffering. He gets the credit for the alleviation of suffering that follows from his efforts. In fact, any end an agent has is something he can be effective or ineffective at bringing about. His effectiveness is both a matter of reliably producing the intended effect and of doing so because of the exercise of his own power rather than because of the many other conditions that are also operative. An effective agent gets the credit for the effect.[8]

One of the ends agents have is to get to the truth. Getting truth is probably the primary epistemic end of agents, or at least a very important end. It is also arguably

a "natural" end of belief formation. Epistemically effective agents therefore have a high proportion of successes in reaching truth and avoiding falsehood. And their successes must be credited to themselves rather than to something else.[9]

Must an agent satisfy counterfactual conditions for effective agency? If so, must she satisfy them on each occasion in which she is exercising agency? Reid maintained that in the case of the active powers (as opposed to the intellective powers), an agent does not have the power to do something unless he has the power not to do it. This is strikingly similar to the Principle of Alternate Possibilities (PAP) often proposed by non-determinists as a condition for moral responsibility: A person is not responsible for her act unless she could have done otherwise. According to PAP a counterfactual condition is necessary for the attribution of moral responsibility to an agent's act. Reid's principle is stronger since he proposes that alternate possibilities are necessary for the power an agent must have to act as an agent, not just for the subset of acts for which he is responsible. Therefore, if PAP fails, Reid's principle fails also.

Thirty years ago Harry Frankfurt presented a famous thought experiment that arguably leads to the conclusion that PAP is false. In this section I want to look at Frankfurt-style cases, but not for the usual purpose; moral responsibility is not the focus of this essay. But the moral of Frankfurt cases, I believe, is important because it can be generalized to apply to many principles that offer counterfactual conditions for the application of some property. Knowledge is often defined in a way that includes counterfactual conditions. Many philosophers who propose such accounts intend these conditions to specify what it takes for the knower to get credit for her belief. Since the idea of credit is similar to that of responsibility, we might expect criteria for responsibility to have an analogue in criteria for epistemic credit, and we might also expect that any problems in the former may be reflected in problems in the latter. Other proponents of counterfactual accounts of knowledge separate epistemic credit from knowledge. But I think that even these accounts can be illuminated by examining epistemic parallels to Frankfurt cases. That is because I believe that Frankfurt has identified a very general problem in counterfactual conditions for any property, whether or not it has anything to do with responsibility. My intention, then, is to see how epistemic Frankfurt-style cases can illuminate both the connection between counterfactual conditions and epistemic credit, and the broader issue of the extent to which counterfactual conditions are necessary for knowledge. My hope is that these thought experiments will make it easier to answer the questions posed at the end of section I.

Case 1: Standard Frankfurt case
Black, an evil neurosurgeon, wishes to see White dead but is unwilling to do the deed himself. Knowing that Mary Jones also despises White and will have a single good opportunity to kill him, Black inserts a mechanism into Jones's brain that enables Black to monitor and to control Jones's neurological activity. If the activity in Jones's brain indicates that she is on the verge of deciding not to kill White when the opportunity arises, Black's mechanism will intervene and cause Jones to decide to commit the murder. On the other hand, if Jones decides to

murder White on her own, the mechanism will not intervene. It will merely monitor but will not affect her neurological function. Now suppose that when the occasion arises, Jones decides to kill White without any "help" from Black's mechanism. In the judgment of Frankfurt and most others, Jones is morally responsible for her act. Nonetheless, she seems to be unable to do otherwise since if she had attempted to do so, she would have been thwarted by Black's device.[10]

Discussion of cases like this has generated a large literature. Non-determinist defenders of PAP have argued that Frankfurt's thought experiment fails to demonstrate the falsehood of PAP,[11] while many determinists have argued that these cases successfully falsify PAP.[12] I have argued that Frankfurt cases are successful in demonstrating the falsehood of PAP, but they are unsuccessful in supporting determinism.[13] Frankfurt cases succeed in showing that principles that offer counterfactual conditions for the application of some property can fail because of the possibility of a counterfactual manipulator, and so the counterfactual condition is not strictly necessary for the application of the property in question. But it would be too hasty to conclude, as Frankfurt does, that the counterfactual condition can fail systematically, much less that it is irrelevant. The reason for this might be that counterfactual conditions are usually proposed not because actual conditions are literally inadequate, but because it aids our understanding of conceptually abstruse properties such as *responsibility, power, causality*, and *knowledge* to think of them in terms of what happens in non-actual circumstances. If I am right about this, counterfactual conditions can fail even when they are relevant and perhaps even when their failure must be selective.

To make the point, let me give a harder Frankfurt case (F case) that changes nothing in the standard case except background conditions.

Case 2: Altered Frankfurt case
In the standard F cases the device is set to go into operation a maximum of once, but it is not needed in the case in question because Jones makes what Black considers the "right" choice on her own. But suppose that Black has been systematically manipulating Jones's choices all along. Every time Jones is about to make a choice, if it is the one Black wants, the device does nothing, but if it is not the one Black wants, the device makes Jones choose the way Black wants her to. And let us suppose that Jones has been living with this device for many years. A multitude of her choices have been manipulated and changed by Black, unknown to Jones. And suppose this is one of those times that Jones makes the choice Black wants and so the device does not go into operation.

Is Jones responsible in this case? Perhaps she is; I am not going to argue that she is not. My point is that the case is harder and it can be made harder still. This could be the only time in her life that Jones has made a choice on her own without the intervention of the machine. If so, we might worry that the counterfactual manipulator is not only manipulating the circumstances; he may be manipulating the person. If we hesitate in ascribing responsibility, I suspect that that is because we think that Jones's lack of responsibility for all her other choices can infect her responsibility for

the choice the one time the device is not needed. A person who never satisfies PAP might be a different kind of being than one who usually does. She is arguably not an agent, the kind of being who *can* be morally responsible. Perhaps moral agency, like virtue, is the sort of thing that takes practice.

But whether or not we ultimately decide that Jones is responsible in case 2, the fact that this case is harder suggests that the counterfactual condition can still be a good sign of the presence of the target property even if it is not necessary that it *ever* be satisfied. If so, what is really essential to responsibility needs to be extracted from a story that explains why we sometimes think the property obtains even when the counterfactual condition is not satisfied. This is important because the wrong moral to draw from these cases is that the counterfactual condition is irrelevant. Frankfurt was right that he had described a case in which alternate possibilities are not necessary for responsibility, but then he leaped to the conclusion that therefore there is nothing blocking the acceptance of determinism. In my view, what Frankfurt cases show is that whereas alternate possibilities are not strictly necessary, they are usually associated with responsibility because they are a sign of something that really *is* necessary—the presence of agency. Agency is a necessary condition for responsibility. If it is lacking, so is responsibility.[14]

If I am right about this, we can apply the same point to properties other than moral responsibility for which there are allegedly counterfactual conditions. Epistemic credit is a property that is closely allied to responsibility and in addition to being interesting in its own right, it may be an ingredient in knowledge. Can we use epistemic Frankfurt-style cases to test the need for something like PAP in cases of epistemic credit and/or knowledge?

> *Case 3: Epistemic Frankfurt case*
> Suppose that Jones is very good at identifying vintages of Bordeaux. In particular, she has no trouble distinguishing a '94 Chateaux Margaux from very similar wines. Black knows that Jones is going to be tasting different vintages of Margaux without knowing in advance the year of the vintage she is tasting. He has installed a device in her head that can make Jones believe that the next wine she tastes is a '94 Margaux whether it is or not. (Never mind why Black would want to do such a thing.) When Jones tastes the next wine, if she appears about to judge that it is a '94 Margaux, the device does nothing. But if Jones is about to judge that it is anything else, the device will interfere with her tasting sensations and will lead her to think it is a '94 Margaux. Now suppose that she tastes a '94 anyway and believes it is a '94, and Black's device does nothing but monitor what is going on in Jones's nervous system. Jones's tasting faculties and taste memory are working fine and she comes to have a true belief in the normal way.

My intuition in this case is that Jones gets epistemic credit, and for the same reason that she is morally responsible in the standard Frankfurt case. Furthermore, I am willing to say that she has knowledge. She knows she is tasting a '94 Margaux. The device does not operate and its very existence is an accidental feature of Jones's epistemic situation. As in the standard F case, the counterfactual manipulator has no effect on our inclination to judge that Jones has the property we are inspecting. But

notice that she not only fails the test of alternate possibilities, she also fails some well-known counterfactual conditions for knowledge. For example, she may fail the Nozick conditions since she would have had the same belief even if it had been false.[15]

We can make up a harder epistemic F case, parallel to the harder regular F case as well.

Case 4: Altered epistemic Frankfurt case
Suppose that it will serve Black's sinister purposes if Jones forms a specific set of beliefs. Some of these beliefs are true, but many of them are false; their truth or falsehood is irrelevant to Black's purposes. Jones's beliefs are systematically manipulated by the device Black has installed in her head. Many times the device has forced Jones to form a false belief that she would not have formed on her own. Other times it forces her to form a true belief that she would not have formed on her own. Still other times it permits her to form a true belief on her own in such a way that in the absence of the device we would not hesitate to say that Jones is credited with getting the truth and knows the proposition in question. Now suppose Jones forms the true belief that she is tasting a '94 Margaux in this way.

My reaction to the altered epistemic F case is the same as to the altered F case. While I do not propose that Jones does not know the identity of the wine she is sipping, I find it a more difficult case. Perhaps believing on her own, like acting on her own, is something that at least in many cases requires a background of practice in believing/acting on her own. Perhaps very simple sorts of perceptual knowledge do not require such a background, and I will have more to say about the simplest perceptual knowledge in section III, but believing out of an acquired power of taste discrimination does seem to be the sort of perceptual knowledge that may require such a background, or it least the possibility that it is required is enough to make some of us worry that Jones does not know in case 4.

Epistemic Frankfurt cases have something in common with the standard evil demon scenarios, but there is an important difference. The F cases involve manipulation of agency in a way that does not appear in the skeptical scenarios. The evil demon gives the agent misleading sensory inputs that inevitably result in false beliefs, but the agent's control over her reasoning process is not altered. This is like deception, whereas the F cases are cases of coercion. Extended manipulaton of the reasoning process itself undermines our ability to initiate our own cognitive projects. The counterfactual manipulator in epistemic F cases therefore attacks the agent's epistemic responsibility, whereas the evil demon does not.[16]

I conclude that the moral of the epistemic F cases is the same as the F cases. Cases 1 and 3 show that the counterfactual condition is not strictly necessary for the target property, but cases 2 and 4 show that it is not irrelevant. It is a sign of something deeper: The agent gets credit for reaching the end. The agent must be an agent, and the fact that she gets a true belief must be due to her. Whether or not her belief is voluntary, her agency is central to acquisition of the belief. Causal processes that bypass her agency take away her epistemic credit, and they also take away her knowledge.

Notice that the application of epistemic F cases to knowledge need not go *through* the concept of epistemic credit. Some accounts of knowledge analyze it in terms of counterfactual conditions, and some of those conditions fail in the epistemic F cases whether or not the reason is that the agent lacks epistemic credit. I have already remarked that case 3 might fail Nozick's conditions even though it is intuitively a case of knowledge, and it no doubt fails the conditions of other theories as well. The conditions for reliabilism are a special and interesting case because they are tied to effectiveness, but not to the "agent gets the credit" aspect of effectiveness. Instead, they are tied to the first condition for effective agency mentioned above—that the agent must have a high proportion of successes over failures. However, the counterfactual manipulator can arguably make the agent unreliable when we think she has knowledge and reliable when we think she does not have knowledge. In case 3 where the device is set to operate a maximum of once, Jones is arguably still reliable and she does have knowledge, as the theory predicts.[17] In case 4 she is unreliable and, as the theory predicts, we hesitate to say she has knowledge because of the worry that the machine has tampered with her agency. But my experience with proposing this case to others leads me to think that the intuitive judgment here is unclear. If Jones is able to form true beliefs in the normal way through the exercise of her own power when the machine is not operating, perhaps it is reasonable to say that she has knowledge even if the machine has made her unreliable.

Furthermore, consider the case of the benign manipulator (Case 5), who makes Jones believe only truths.[18] In such a case Jones is both highly reliable and satisfies some counterfactual conditions for knowledge, but it is doubtful that she has knowledge when she acquires a machine-produced true belief.[19] An agent reliabilist who shares this intuition might say that the problem here is that *Jones* is not reliable. What is reliable is the machine operating in her. If so, the problem I am raising is not a problem for agent reliabilism itself, but for the view that reliability entails counterfactual conditions that are subject to Frankfurt-style manipulation. A careful agent reliabilist could therefore accommodate the intuition I have about case 5 and perhaps case 4 as well. In the latter case the agent reliabilist could say that the agent herself is not unreliable. What is unreliable is the complex of agent plus machine.

Similarly, the process reliabilist could say that what is unreliable in case 4 is the process used by the complex of agent plus machine. In case 5, if we assume that the benign manipulator does not generate the belief directly, but makes the process the agent is using or her faculty reliable, there is no problem in concluding that she has knowledge. After all, even ordinary knowledge often makes use of aids to our faculties: eyeglasses, hearing aids, and one day, perhaps, computer chips installed in our brains to aid our memory. These devices make it easier for us to get knowledge; they surely do not take it away. So if the benign manipulator in case 5 installs a device that is comparable to these aids, only better, there is no difficulty. But there are ways the benign manipulator could operate that would threaten Jones's epistemic agency or even eliminate it. The device could bypass Jones's perceptual and cognitive functions entirely. Suppose it implants true beliefs in her head overnight, counteracting any false beliefs she unhappily acquired during the day and adding many

other true beliefs besides. The process is as reliable as a human-generated process can be. Does Jones have knowledge?

Here I think our agreement on the target property we are trying to analyze breaks down. Some philosophers will probably be willing to say that Jones has knowledge in this amendation of case 5. But if so, they are not treating knowledge as something we earn or even something we contribute to through our own powers; it is something we are blessed with. I recognize that it is not obvious that knowledge is a unitary concept, and it may be a flexible enough concept to include instances of knowledge as a gift rather than as something we merit. But I think we should try to see how far we can get with a unitary concept, and if we are going to aim for a unitary concept, it should be one of merit, not blessedness. If so, Jones does not have knowledge when her true beliefs are produced wholly through the action of a benign manipulator.

I conclude that we should be wary of making manipulable counterfactual conditions necessary for either epistemic credit or knowledge (case 3), nor are they sufficient for knowledge (case 5). But neither should we ignore the importance of the close association that ordinarily obtains between those conditions and knowledge.

III. Knowledge, Agency, and Virtue

Let us now go back to giving a partial answer to the three sets of questions posed at the end of section I. I have already answered the general question of what it takes to be an effective agent. An effective agent is reliably successful in reaching her ends and she does so through the exercise of her own power. Frankfurt-style cases show that efficacy is associated with the satisfaction of counterfactuals, but cases 1 and 3 show that their satisfaction is not necessary on every occasion in which agency is operative, and case 5 shows that it is not sufficient. The satisfaction of appropriate counterfactuals is not constitutive of agency, but is a sign of it.

For the same reason, it is not strictly necessary that I be reliably effective in order to effectively exercise agency on a particular occasion. In cases 2 and 4 the agent is unreliable but she is arguably successful in exercising her agency, although some of us hesitate. Our hesitation, however, does not stem from her lack of reliability per se, but from the worry that the counterfactual manipulator has interfered with her power to be an agent. And case 5 shows that reliability is not sufficient. Again, I think that the fact that an act/belief is that of a reliable agent is a sign of what we are looking for—that the act/belief really belongs to the agent; that she gets credit for it.

What difference does it make to epistemology if a causal process is brought about by the exercise of agency? Since getting the truth is one of my ends, I am an effective epistemic agent to the extent that I am reliably successful in reaching the truth and do so because of the exercise of my epistemic powers. My epistemic success is due to me. In case 5 the benign manipulator makes Jones believe only truths, and we considered a causal process that bypasses Jones's perceptual and cognitive faculties completely. My intuition is that Jones neither gets epistemic credit nor has knowledge

in this case. So it matters epistemically that her agency is not operative. It also matters epistemically that her agency *is* operative in cases 3 and 4. In case 4 the intuition that she knows and gets epistemic credit is weaker than in case 3, but that is because of doubts about her agency. If she really is acting as an agent, the intuition that she knows is fairly strong even though it is clear that she is unreliable. Agency seems to be enough to make up for the lack of reliability, whereas even the presence of reliability cannot make up for the lack of agency.

In stressing the importance of agency in getting epistemic credit and knowledge, it must be admitted that human agents are not pure agents and it is unreasonable to expect otherwise. Our beliefs, like our desires, often come unexpectedly as the result of causal processes that are largely external to ourselves, and there is nothing abnormal about that. Many philosophers are willing to say that some of these beliefs constitute knowledge. In particular, some perceptual knowledge may be in this category. The examples of acquired wine discrimination show that not all perceptual beliefs can be in this category, and probably most perceptual beliefs utilize some degree of learned discrimination. But perhaps the simplest cases of perception do not require the operation of agency. What should we say about these cases?

I have argued elsewhere that agency operates counterfactually even in simple perceptual cases: If an intellectually virtuous agent had indications that her perceptual ability or her perceptual situation was in some way deviant, she would withhold or withdraw perceptual judgment until she could investigate.[20] If this is right, agency operates even in the simple cases of automatic perceptual belief formation at the second-order level, the level of reflectiveness. But let us look once again at the Frankfurt cases since I think we can use them to illuminate the place of agency in evaluating simple perceptual beliefs.

So far I have said nothing about what happens from the agent's viewpoint when the Frankfurt device operates. Although we can only guess at the phenomenology of the device during operation, I imagine that the victim has the experience of impulsively deciding/believing something unexpected. She is about to decide not to kill White when she suddenly decides to kill him after all. Or she is about to believe that the wine she is sipping is a '95 when she abruptly decides it's a '94 instead. Since all of us are subject to changes of mind and sudden impulses, this will not necessarily seem peculiar unless it happens very often or if the decision/belief seems to the agent to be out of character, something she can hardly imagine herself doing/believing. But when an agent suddenly acquires a belief or suddenly makes a decision, she should subsequently reflect about her own belief/decision. I think that agency requires this. One of the central features of agency is self-reflectiveness, and since one of the aspects of self-reflectiveness is the second-order desire for self-integration, agents need to tell themselves *some* story about the unexpected act or belief. I am not suggesting that this is something we should do constantly, and certainly not obsessively. But at some point we should assess our sudden beliefs and decisions, at least those that have any important consequences or implications for our view of ourselves. A sudden belief that comes out of nowhere, like a sudden urge, ought to be either endorsed or repudiated. Agency does not require that we do one rather than the other, but it does require that we do one or the other, probably not for every sud-

den belief and impetuous act, but certainly for some of them. I suggest that when the machine operates in cases 2 and 4, a test of whether the machine has eliminated Jones's agency is whether she reflectively endorses or repudiates her machine-produced beliefs/acts after the fact.

But, you will ask, cannot the device also operate on the second level, the level of endorsement? Yes, of course it can, and if it does, and if there is no higher level of reflectiveness at which the machine does not operate, then it is likely that Jones has indeed lost her agency.

In some respects our simplest perceptual beliefs are like the beliefs produced by the counterfactual manipulator. They come upon us without warning and without any effort on our part. In normal situations there is nothing suspicious about them and we have learned to expect to have perceptual impressions almost all the time, so the beliefs formed from these impressions are not like suddenly acquiring the urge to kill. If I am normal, these beliefs are easy to integrate into my view of myself and my environment, unlike the urge to kill or the sudden belief that airplanes are following me. Perceptual beliefs are typically unimportant, and there is no great need for reflective endorsement in many cases. But when the consequences of believing them are serious, reflective endorsement is called for. If Jones is a professional wine taster she should reflect about the grounds of her belief when tasting wine, at least when something of importance hinges on her judgment. If some of her beliefs are unknowingly machine-produced, she should reflect about them if she has reason to suspect that there is something out of the ordinary in the way she got them. Since reflectiveness preserves her agency, she can get epistemic credit even when the belief is machine-produced. It is even possible that she gets credit for her beliefs in case 5 if she later endorses them when the machine is not operating. For the same reason, she can get epistemic credit for non-voluntary perceptual beliefs. An agent reflects about her beliefs from time to time, particularly when they are either suspicious in their origin or of special importance. And this includes perceptual beliefs. True perceptual beliefs earn the believer epistemic credit when the agent exercises her agency over them at the level of reflective endorsement.

The view I am proposing on the place of agency in belief is similar to Christine Korsgaard's interpretation of autonomy in desire. Korsgaard argues that according to Kant, autonomy is compatible with acting out of desire as long as the reflective mind endorses the bidding of desire. In this way we are self-determining even when we act instinctually.[21] Similarly, I am suggesting that we can be autonomous agents even in the simplest perceptual knowledge by endorsing the bidding of our pre-reflective minds. The connection between the reflective endorsement of belief and the second-order endorsement of desire has been explored in some detail by Keith Lehrer. Lehrer calls the positive evaluation or endorsement of desire "preference" and the positive evaluation or endorsement of belief "acceptance."[22] I am not suggesting that the place of agency in knowledge and responsible belief is limited to such second-order endorsements, but it is a way in which agency can extend even to those parts of the self that are initially acquired non-voluntarily. Since it is likely that some perceptual beliefs — those requiring the most meager conceptual resources — are in this category, it means that agency can extend even to such perceptual beliefs.

I suspect, then, that self-determination, autonomy, and agency operate on a much wider scale than is included in intentional action or even the broader class of voluntary action. If I am right about this, not only is it a mistake to focus on the voluntary/non-voluntary distinction in analyzing epistemic responsibility, it is a mistake in the analysis of moral responsibility as well.

This position obviously needs considerable refinement and a number of objections need to be answered. The scope of the self is a difficult matter, and the Kantian view of the self is notoriously narrow. Why think that our first-order desires and beliefs are any less a part of our selves than our will or second-order endorsements? Why does the latter have authority over the former? It is far too facile to identify the self only with those desires/beliefs with which we identify at the second level. After all, a person who has numerous first order desires or beliefs with which he does not identify has a different self than he would have had if he had not had them. He has, we would say, a fragmented self. And there may even be a sense in which he is responsible for the desires/beliefs with which he does not identify or even explicitly repudiates. That might explain why even such desires/beliefs are in some sense his own. Some epistemologists wish to extend the scope of knowledge to include many such beliefs. I've already said that knowledge may not be a unitary concept and there may not be any way to resolve some of the disputes about the application of "knowledge" to cases in which agency clearly does not apply, but I am suggesting that we need not assume that agency does not apply when an epistemic state is initially acquired non-voluntarily. Furthermore, I suggest that part of the reason it is hard to decide whether knowledge or epistemic responsibility applies to beliefs about which agency does not apply is the vagueness of the boundaries of the self.

We can now answer the second question posed at the end of Part I: Is there any important difference between an effect arising from the act of an agent, whether voluntary or non-voluntary, and events brought about by a non-agent? In particular, does it make any significant difference to epistemology? We have seen that the non-voluntary acts and beliefs of agents can differ in important ways from events that are produced by non-agents. If it is the act or belief of an agent, the agent's subsequent reflectiveness makes it voluntary on the second level. The agent either does or does not make the belief her own.[23] Even non-voluntary acts/beliefs can therefore earn the agent credit (or blame), and in the case of beliefs, they may constitute knowledge.

This brings us to question (3): Is knowledge best understood on the model of event causation or on the model of agent causation? I have already suggested that epistemic credit is earned by an agent only when her agency is operative, either in the initial acquisition of the belief, or in her later reflective endorsement of the belief or beliefs like it. Since on my view epistemic credit is a component of knowledge, I am also willing to say that knowledge requires the operation of agency. But throughout this essay I have not rested my case that the agent does or does not have knowledge in the various Frankfurt cases on the fact that she does or does not have epistemic credit. In case 3, for example, it seems to me that the agent has knowledge. It also seems to me that she gets epistemic credit. But I am not suggesting that she

has knowledge *because* she gets credit. Those epistemologists who separate epistemic credit, either in the form of justifiedness or responsibility, from knowledge will no doubt have more complicated responses to these cases. They might say, for instance, that agent causation need not be operative in knowledge, but it does need to be operative in generating justified or responsible belief. But I will not try to sort out here the various possibilities that emerge when various forms of epistemic credit are separated from knowledge. My position is that the fact that an agent has knowledge is "up to her," to use Reid's words. She need not be responsible in the sense that requires alternate possibilities or even voluntariness, but she needs to be exercising her agency either at the first- or second-order level.

Aquinas defines virtue as the perfection of a power.[24] Within the context of this essay that would mean that virtue is a property that makes agents effective. In a broad sense of virtue there can be physical virtues, culinary virtues, teaching virtues, and so on. Epistemic virtues make us effective epistemic agents. An effective epistemic agent is one who reliably reaches her epistemic end and who reaches her epistemic end because of *her*, not by chance or because of something outside of her. Need she be exercising a virtue in getting knowledge? I have said elsewhere that that is too strong a requirement.[25] She does need to be exercising a power and she needs to get to her end because of her power; she must be exercising her agency. She need not intentionally aim at her end, however. She need not even be acting voluntarily. In fact, it might turn out that she need not even be generally reliable in reaching her end, assuming that in case 4 we ultimately judge that she knows. But she needs to be an agent. Her agency is critical in explaining how it is that she ends up with her true belief. A causal sequence that leaves out her agency is not good enough.

At the beginning of this essay I remarked that ethics is concerned with the evaluation of agents and their acts, but that does not tell us the full range of ethics since both agency and act could be more or less extensive than we think. I have argued that agency is operative in getting epistemic credit and knowledge. The scope of agency includes those evaluative aspects of belief investigated by epistemology. In other work I have argued that it is artificial to separate epistemology from ethics. The role of agency in beliefs as well as in acts further supports this position.

Notes

1. To complicate matters further, it is likely that not everything an agent "does" is an act, although it is difficult to draw a systematic distinction between a non-voluntary act and a non-act done by an agent. I will not pursue this distinction here. See Jonathan Bennett, *The Act Itself* (Oxford: Clarendon Press, 1995), beginning of chapter 2.

2. Agent causation is sometimes even thought to be necessary to explain event causation since the latter generates a regress ending in a non-event: an agent.

3. Susan Sauvé Meyer denies that Aristotle contrasted agent causation with event causation since on Aristotle's view, every efficient cause is a substance, whether or not it is an agent and, in fact, the effect is a substance also. See "Self Movement and External Causation," in *Self-Motion*, Mary Louise Gill and James Lennox, eds. (Princeton, N.J.: Princeton University Press, 1994). But for the purposes of this essay it is worth pointing out that whether or not Aristotle had a notion of event causation, his notion of efficient causation

was more like what we mean by agent causation than event causation. The causation due to human agents is a subclass of the substance causation found in nature. A broader definition of agent causation is given by William Rowe, who defines "agent causation" as "the idea that the primary cause of an event is a substance." Rowe claims that Reid used "agent causation" in a narrower sense (*The Cambridge Dictionary of Philosophy*, Robert Audi, ed., 1995, p. 13). On Rowe's definition, Aristotle's notion of causation would qualify as agent causation.

4. See Reid, "Of the Liberty of Moral Agents," chap. 2, and "Of the Words Cause and Effect, Action and Active Power," in *Essays on the Active Powers of the Human Mind,* (Cambridge, Mass.: MIT Press, 1969), 267ff.

5. Not only did Aquinas place a great deal of importance on the cognitive act, but Eleonore Stump argues that he believed most acts of intellect are not causally determined. See "Aquinas's Account of the Mechanisms of Intellective Cognition," *Revue Internationale de Philosophie* 21 (1998): 287–307.

6. There is more than one sense of ends used in the history of ethics and sometimes they are not clearly distinguished. For the purposes of this essay, ends can be those at which the agent consciously aims, or they can be natural ends. Either interpretation is permissible.

7. These two senses of chance are distinguished by Wayne Riggs in "What are the 'Chances' of Being Justified?" *The Monist* 81, 3 (July 1998): 452–72.

8. Of course, no effect is brought about by a single cause, whether the cause is an event or an agent exercising a power. But some causes are much more salient than others, and that leads us in many of the most interesting situations to designate one or a small number of causes as "the" cause. As far as I can see, the vagueness of this usage does not affect the argument of this essay.

9. Wayne Riggs addresses the issue of the need for the knower to get get credit for her belief in "Reliability and the Value of Knowledge," forthcoming, *Philosophy and Phenomenological Research*.

10. This adaptation of Frankfurt's example using a neurological device is similar to some of the cases described by John Martin Fischer. An early use of this type of example appears in "Responsibility and Control," *Journal of Philosophy* 89 (January 1982): 24–40.

11. Some of the recent essays taking this position are David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternate Possibilities," *Philosophical Review* 104 (April 1995); Robert Kane, *The Significance of Free Will,* (New York: Oxford University Press, 1996); Michael McKenna, "Alternate Possibilities and the Failure of the Counterexample Strategy," *Journal of Social Philosophy*, 28, 3 (Winter 1997); and Michael Otsuka, "Incompatibilism and the Avoidability of Blame," *Ethics* 108 (July 1998): 685–701.

12. The most extensive deterministic defense of the success of Frankfurt's counterexamples to PAP has been given in a number of places by John Martin Fischer. See *Metaphysics of Free Will,* (Cambridge, Mass.: Blackwell, 1994), chapter 7.

13. I first argued for this position in *The Dilemma of Freedom and Foreknowledge* (New York: Oxford University Press, 1991) and most recently in "Does Libertarian Freedom Require Alternate Possibilities?," in *Philosophical Perspectives*, vol. 14, edited by James Tomberlin, 2000. The same position has been argued by Eleonore Stump in "Intellect, Will, and the Principle of Alternate Possibilities" in *Christian Theism and the Problems of Philosophy*, Michael D. Beaty, ed. (University of Notre Dame Press, 1990); by David Hunt in several places, most recently "Moral Responsibility and Unavoidable Action," forthcoming in *Philosophical Topics*; and by Derk Pereboom in "Alternative Possibilities and Causal Histories," forthcoming in *Philosophical Perspectives*.

14. It would be fair to point out that in drawing this conclusion I have moved backward in the order of explanation. Agency is a vague concept, just as vague as responsibility. It does not help us understand what responsibility is to be told that it requires agency. In contrast,

PAP at least has the virtue of being clear and reasonably precise. I agree with this point, but deny that it is an objection. I am not offering any part of an account of responsibility here. My point is that we should not worry excessively about the failure of PAP. Its failure permits us to look more deeply at the property PAP was aiming to elucidate.

15. Perhaps she does not fail the Nozick conditions since he requires that the method of belief formation must be kept constant. She might, therefore, satisfy the following condition: If the belief *p* had been false *and she had used the same way of arriving at whether* p, she would not have believed that *p*. Arguably, in the epistemic F case, if *p* had been false she would not have used the same way of arriving at whether *p*.

16. I thank Abrol Fairweather for this point.

17. In fact, she may not be reliable on some ways of construing reliability, in which case our basic epistemic F case is a counterexample to such theories.

18. Of course, unless the benign manipulator is omniscient, he will not be able to fully accomplish this, but we need only assume that the benign manipulator has far greater knowledge than Jones.

19. Compare what we would say about the parallel moral case. If the benevolent manipulator makes Jones do only right acts, a right machine–produced act does not earn Jones any moral credit.

20. Linda Zagzebski, *Virtues of the Mind*, (Cambridge: Cambridge University Press, 1996), 280, and in the "Reply to Alston" in the Symposium on *Virtues of the Mind, Philosophy and Phenomenological Research*. Careful readers will no doubt notice that this is a counterfactual condition and perhaps it also can fail due to the action of a counterfactual manipulator.

21. Korsgaard discusses reflective endorsement in a number of places in *Sources of Normativity* (Cambridge: Cambridge University Press, 1996). See in particular the section entitled "The Status of Desire," in her reply to Geuss, 238–42.

22. See Keith Lehrer, *Self-Trust* (Oxford: Clarendon Press, 1997).

23. This naturally raises the question of whether alternate possibilities are necessary at the second level. I will leave that question aside for this essay.

24. *Summa Theologiae* I-II, q. 55.

25. Zagzebski, *Virtues of the Mind*, Part III, sec. 2.

10

EPISTEMIC LUCK IN LIGHT
OF THE VIRTUES

Guy Axtell

> The problem of restoring integration and cooperation between man's
> beliefs about the world in which he lives and his beliefs about the
> values and purposes that should direct his conduct is the deepest
> problem of modern life. It is the problem of any philosophy that is not
> isolated from that life.
>
> — John Dewey

I. Luck, Control, and Virtue Theory

In its broadest context, as part of contemporary virtue theory, virtue epistemology can be understood as one side of an integrated account of the human being as both a knower and a valuer. What attracts me to it, personally, is my sense of the fragmented state of reason in contemporary western culture, a condition that is to some extent an inheritance of modern philosophy. Virtue theory promises a response to what Dewey, quoted above from *The Reconstruction of Philosophy,* presents as a central problem of modern life and even as *the* theoretical chasm for philosophy to bridge: the chasm between our beliefs about the world and our beliefs about the virtues, rules, and purposes that should direct our actions, both individually and collectively. I use these metaphors of bridge and chasm in connection with Dewey's passage, because this is an essay *about* bridges: bridges, firstly, between epistemology and ethics; and secondly, between virtue epistemologists themselves since they, like others, invariably bring with them to the discussion of knowledge their own divergent interests in explanation.

Our approach will be through an analysis of the various possible meanings of "epistemic luck" and their impact on epistemology. Luck presents a complex of problems that impact both ethics and epistemology, problems that must be balanced with and related to our conceptions of responsible agency. The presence of luck in

our cognitive as in our moral lives shows that the quality of our intellectual character may not be entirely up to us as individuals, and that our motivation and even our ability to desire the truth, like our moral goodness, is fragile. But the fragility of our character makes it a more, not a less, important topic for philosophical inquiry. As Claudia Card and Martha Nussbaum have both pointed out, the impact of luck on our lives can also add depth to our understanding of responsibility and increase our sense of the worth of the virtues.[1]

Luck also affects our understanding of the reliability of our belief-forming cognitive processes (hereafter BCPs). The reliability of these processes is in one sense radically subject to luck, to a "natural lottery" as some would describe it, or to the control not of ourselves but of nature. To the extent that our intellectual habits and dispositions can be influenced by choices under our control, however, their reliability might be taken as an achievement. But to what extent is this? There is keen debate among epistemologists about whether our intellectual dispositions—those that affect our BCPs—are less subject to voluntary control than are the moral dispositions that shape our actions. Views on this matter are bound to impact our conceptions of both ethical and epistemic agency. It is a commonplace to hold that the more externalist an epistemologist's orientation, the more tolerance they have for luck in our cognitive lives.

Here is a first indication of differences among virtue epistemologists: those inclined toward reliabilist views tend on the whole to be more skeptical concerning issues of responsibility for one's intellectual character than those inclined toward what has recently come to be called "responsibilist" views.[2] For instance, in a recent essay, "Moral and Epistemic Virtue," Julia Driver, who takes a stance she calls objective consequentialism[3] allied with "the externalist extreme" of a thoroughgoing reliabilism, writes that "Luck is a fact of life, and it may be a form of moral and intellectual virtue to humbly accept that there will be such limits to success no matter how well the agent is justified." Those who think epistemic agents have little control, or only very indirect control over their intellectual dispositions and habits, lean toward reliabilism even if, as Linda Zagzebski has suggested, there may be few necessary connections to be found between externality and lack of control, or between internality (awareness) and control.[4]

The issue of *responsibility for character* (which would appear to follow from our capacity to control or affect our own traits) is a difficult one, and all the more so when we extend the term "character" to include intellectual habits and dispositions in addition to moral ones. The analogies between ethics and epistemology are undoubtedly on stronger ground when one compares the degree of control we enjoy over ethical and intellectual *character traits* than when one attempts to compare our control over our *actions* with our control over our *beliefs*. But terms of character such as "disposition" and "habit" are used to cover a wide range of attributes not well demarcated by their openness to conscious or to voluntary control. Moreover, epistemologists often invoke different vocabularies when issues arise about control. Some speak without worry about a "weak doxastic voluntarism," from which control and hence responsibility flows, while others speak as if epistemic agents are about equally responsible or non-responsible for each of their beliefs (certainly not something we

hear said about each of our actions!). Both vocabularies can contribute to conflation of the descriptive and evaluative tasks of epistemology. The former is odd if it assumes one *needs* to support doxastic voluntarism in order to speak about control over traits of intellectual character. The latter is odd in light of the generally accepted points that "*S* believes *p*" is attributable to an agent *S* (with the possible exception of perceptual beliefs) only on the basis of a prior assumption that the agent *S* does (or would) accept the proposition that *p*, and that the "justification" for acts of acceptance is not an all-or-nothing affair, but rather comes in degrees.

Hilary Kornblith avoids this conflation between the normative and the descriptive by distinguishing between "actions" and "processes," that is, between actions that we *ought* to perform in order to make the processes by which we arrive at our beliefs more conducive to truth, from those processes (BCPs) themselves. A related suggestion, to which I will try to adhere, is that in order to demarcate the proper application of normative concepts of agency like "rational," "responsible," "praiseworthy," etc., we should distinguish between the *acceptance of a proposition* and the *state of belief*. Whichever distinction is preferred, formulating criteria for evaluating an agent's *acceptance* of a proposition involves these normative concepts of rationality and responsibility.

In general I find differences over the issue of control over intellectual habits and dispositions to run less deep among virtue epistemologists than among epistemologists at large. Part of the reason for this may be that virtue epistemologists share a broad range of assumptions that bring them together with interest in issues—both the descriptive and evaluative—affecting our understanding of epistemic agency. Moreover, one opinion virtue epistemologists widely hold is the present need for a "dual component" conception of knowledge, sometimes also called a "mixed" externalist account of knowledge.[5] This is one that integrates constraints on an agent's faculty reliability with constraints on the agent's responsibility in gathering and processing evidence. Kornblith divides the latter into "internal coherence" (desire for epistemic integration) and "action-theoretic" constraints (characterized in terms of desire for and effort in attaining the truth). To briefly cite three further instances, Ernest Sosa writes that in his virtue perspectivism, a proposition is evident or known (from the *K* point of view) to a subject "only if he is *both* rationally justified in believing it *and* is in a position to know (from *K*'s point of view) whether it is true."[6] In her *Virtues of the Mind* (1996), Linda Zagzebski, uses the term "dual-component" to describe her mixed account of *believing from virtue*, which includes both a motivational and a success component. And John Greco says that his account is intended to satisfy both the "subjective justification" and the "no accident" conditions on knowledge by drawing on the resources of virtue theory.[7]

Mixed accounts, I think, present distinct advantages over their competitors because they insist on the complementarity of two concerns that have driven recent developments in epistemology: what it means to have beliefs formed by reliable cognitive processes, and what it means for agents to be responsible and well-motivated in their acquired cognitive habits and dispositions. But differences persist among virtue epistemologists as elsewhere in philosophy and can at times be quite pronounced. Ernest Sosa, John Greco, and Alvin Goldman describe their

virtue epistemologies as specific forms of reliabilism. Lorraine Code formulated the distinction and contrast between reliabilist and responsibilist virtue epistemology in her *Epistemic Responsibility* (1987), arguing that "the concept 'responsibility' can allow emphasis upon the active nature of knowers/believers, whereas the concept 'reliability' cannot."[8] I will draw a distinction, as I have elsewhere, between those inclined toward *virtue reliabilism* and those inclined toward *virtue responsibilism*.[9] The differences that these labels identify should not be exaggerated, since these two hardly represent anything like the extremes of externalism and internalism of the recent past in epistemology. Indeed the issues of first-person constraints such as "internal access" to the grounds of one's belief seems not to be the dividing line between virtue reliabilist and responsibilist accounts of justification. Neither thinks of the nature of justification as turning on that issue, so closely associated with the received definitions of internalism and externalism.[10] But I will take the descriptions of an author as a virtue responsibilist or virtue reliabilist to have a useful application in pinpointing differences of *interest* and *attitude,* and to often be applicable to particular authors even where a "mixed" account may be explicitly advocated.

I hope to show how an analysis of epistemic luck can lend substantial support to mixed externalist epistemologies. But the analysis of luck that will be offered is unique, and first I want to show how problems that surround epistemic luck have in fact served to *divide* reliabilist and responsibilist accounts in the recent past. If we concern ourselves first with the issue of the role of the will and responsibility in propositional acceptance, we find that it is the responsibilists who are most keen to explain the ground for evaluative judgments of epistemic agents: They remind us, for instance, that the Aristotelian account (which allows attributions of responsibility for intellectual habits and dispositions) is not committed to a strong voluntarism (Hookway); that only weak voluntariness is at issue in responsibility (Montmarquet)[11]; and that that there is a *range* of voluntariness in belief as in action—where "accident and intentional action are two ends of a spectrum of conscious control" (Zagzebski).[12]

Reliabilists for their own part are likely to point out, as does Driver, that "there is a built-in responsibility for action" that does not seem to be paralleled in epistemology, where she believes we are "more forgiving" of defects. But this contrasting emphasis is by no means to reject responsibility for character *tout court*, nor by itself does it signal any deep division between our two sub-groups within virtue epistemology. Underlying the mixed account of knowledge is what Thomas Nagel would term a "compatibilist account" of cognitive freedom and causality. This is the epistemic analogue of the view in ethics that one may be responsible for what one does, even if what one does depends in important ways on factors not within one's direct control. In epistemology as in ethics, Nagel tells us, compatabilism "would leave room for the ordinary conditions of responsibility—the absence of coercion, ignorance, or involuntary [actions]" without excluding from the analysis all influence of factors that are a matter of luck.

Let me clarify this last point. In his influential essay "Moral Luck," Nagel frames the relevant epistemic analogue to his compatibilist position in ethics:

The corresponding position in epistemology would be that knowledge consists of true beliefs formed in certain ways, and that it does not require all aspects of the process to be under the knower's control, actually or potentially. Both the correctness of these beliefs and the process by which they are arrived at would therefore be importantly subject to luck. The Nobel Prize is not awarded to people who turn out to be wrong, no matter how brilliant their reasoning.[13]

Such a compatibilist stance in epistemology, to continue the analogy, is simply one that would leave room for the ordinary conditions of epistemic responsibility—primarily the *absence of coercion*—without excluding all influence of factors that are a matter of epistemic luck. Yet this shared compatibilism aside, we can certainly identify issues where the reliabilist and the responsibilist are more sharply divided. Let us examine these. One such issue is how far the analogy between ethical and epistemic luck can or should be pushed. Zagzebski has commented that the recent interest in epistemic luck "makes the attempt to model epistemic evaluation on moral evaluation easier to do," because it indicates that "it is much too facile to distinguish evaluation in the two areas on the ground that we control the one but not the other."[14] Our intellectual habits and dispositions, she holds, are generally not less subject to voluntary control than are their ethical counterparts, although again a *range* of voluntariness must be acknowledged in regard to both. Not everyone would share these responsibilist views, and I use them to illustrate that it is much easier to show how the theoretical differences between virtue reliabilists and responsibilists are manifested than it is to locate and sort out the interests in explanation that drive their divergent epistemological accounts.

In the most widely discussed example, overt differences are manifested in sharply divergent definitions of the intellectual virtues themselves. Reliabilists define them by their conduciveness to the production of true beliefs, allowing genetically endowed faculties and powers to be counted among the virtues; responsibilists tend to take them more restrictively in terms of *acquired* habits and dispositions—traits internal to agency that are the proper object of praise and blame. Responsibilists may also reject the attempt to define the virtues consequentially by what they *do*, in favor of viewing them as conceptually prior to and as partly *constitutive* of the epistemic goal itself.[15]

The differences between reliabilism and responsibilism manifest in another way of which we should take notice: in debate over the possibility and desirability of a *unified account* of ethical and epistemic virtue. Julia Driver usefully distinguishes between the claim that an *account* of the virtues is unified and the claim that the virtues *themselves* are unified. What sense should we give to the latter, more contentious claim? If it is understood as the robust classic Socratic/Aristotelian claim that a person cannot have the intellectual virtues without the moral virtues (and conversely), then it probably has few supporters.[16] But there are various senses of a unified *account* of the virtues that many responsibilists, and only responsibilists, want to defend. The strongest of these may be the *subsumption thesis* that Zagzebski argues for in *Virtues of the Mind* (1996), where she "subsumes the intellectual virtues under the general category of the moral virtues, or *aretai ethikai*, roughly as Aristotle understands the latter."[17]

Less strong, Montmarquet provides an account of the relationship between the ethical and epistemic virtues that is critical, on the one hand, of Zagzebski's strong "assimilationist position," and on the other, of the strong "externalist anti-assimilationist position" which Driver may be thought to hold. The model Montmarquet develops for the relationship between the virtues does not arrange them by content or domain, but rather by the susceptibility of each disposition to *direct control*. Central to his understanding of control is the agent's efficacy in exemplifying a particular virtue at will, i.e., where "trying" to exemplify the virtue is sufficient for exemplifying it. This innovative approach issues in a model reflecting a threefold distinction and "a hierarchy of evaluatively relevant qualities of persons, which largely cuts across the supposed epistemic/moral divide":

> At the lowest level will be those "externalist" traits subject only to our indirect control. At the middle level will be the "internalist" epistemic virtues and perhaps certain broadly moral virtues. At the summit, will be certain narrowly moral traits [picked out by the fact that we are able both to exemplify them and to exemplify their contraries at will].[18]

A unified account of the virtues may be associated with a still-weaker thesis, but one that like the preceding claims also invites "cutting across" some of the disciplinary boundaries between ethics and epistemology. This thesis is one that associates virtue epistemology with the development of a *general theory of value*. Montmarquet and Zagzebski have both supported a sense of this thesis in their book-length treatments of virtue epistemology. I have as well, in previous essays that connect the appeal of virtue epistemology with a broader metaphilosophical reorientation in philosophy. Responsibilists, then, side with Zagzebski's general claim in *Virtues of the Mind* that "the unification of moral and epistemic evaluation is a welcome advantage." They also tend to resist the drift of reliabilism toward a view of epistemology as a disparate field that studies various but unrelated ways in which beliefs can be justified, and thereby side with Paul Bloomfield's general claim that "it would be preferable to find 'justification' to be helpfully univocal."[19] By contrast, those of reliabilist orientation tend to be more sanguine about the prospects of unification in any of the senses we have discussed, and so we find Driver arguing "that a unified account of virtue is neither doable nor desirable."[20]

Let me give just one substantial example of how this debate plays out, using Zagzebski and Driver to exemplify sharply divided responsibilist and reliabilist viewpoints. Zagzebski rejects any sharp distinction between moral and other sorts of evaluation, which she thinks is a residue of the faulty Kantian view that the *good will* (seen as radically free in its *noumenal* existence) is the only proper object of moral evaluation. But her reliabilist critic may perceive her [in a sense analogous to Bernard Williams's critical perspective in *Moral Luck* (1981)], as actually involved in a neo-Kantian attempt to immunize epistemology from luck by exaggerating the freedom and control of the agent (in this case the epistemic agent). While Zagzebski, like Williams, criticizes the Kantian attempt to ground moral evaluation in something that is luck-free, she also takes it as a definite advantage of an epistemic

theory that it "reduces the component of luck." Taking the reduction of the impact of luck as an advantage may seem unobjectionable, but in the context of Zagzebski's approach it reflects her attempt to develop a "motivation-based virtue theory" that places goodness within motivational structure and identifies certain motives as intrinsically good. Seeing Zagzebski's responsibilist account in light of this project makes it difficult to separate Driver's complaint that a motivation-based virtue theory "doesn't provide a criterion" and leaves "mysterious" what makes a motive good or bad, from Williams's notable objection addressed to the Kantian, that "the dispositions of morality, however far back they are placed in the direction of motive and intention, are as 'conditioned' as anything else."[21]

To summarize, we have seen that virtue epistemologists often explicitly disagree over the understanding and identification of the virtues, over the issues of responsibility for character, over the strength of the analogy between ethical and epistemic evaluation, over the prospects for a unified account of the virtues, and over the possibility of a univocal sense of epistemic justification. But if these issues, as we said, merely mark spots where the pot boils over — where explicit theoretical differences persist and main lines of division are manifested among virtue epistemologists — then we should now be enticed to look further and to seek the source of these differences in their divergent interests in explanation. The problem of epistemic luck is an especially promising area to explore in seeking these deeper sources.

II. Forging Bridges: Riggs's Conjunctive Approach

Is knowledge compatible with epistemic luck? An initial step in answering this fundamental query is to distinguish questions about the *extent* of epistemic luck from questions about the *kinds* of epistemic luck thought to have an impact on human cognition. Luck in all its senses appears to involve discontinuities or lack of control either over a process or over its outcome. And as Nagel points out, "epistemological skepticism arises from consideration of the respects in which our beliefs and their relation to reality depend on factors beyond our control."[22] Epistemic luck is a generic notion, and some kinds of luck may simply be irrelevant to an account of justification. But it is a highly contested issue which types present a relevant challenge to knowledge claims. Both groups of virtue epistemologists concede that we cannot always avoid error without luck, but they typically see the relevant discontinuity, and hence the *kind* of luck involved, very differently.

In an innovative recent essay, "What Are the Chances of Being Justified?" Wayne Riggs argues "that both truth-conducivist and responsibilist conceptions of epistemic justification are directed at disallowing a particular kind of chance (true) belief from counting as epistemically sanctioned."[23] Many of Riggs's examples are drawn from virtue epistemologists, making his analysis especially pertinent here. "Chance" is selected as the common term characterizing conditions that might preclude instances of true belief from constituting knowledge, and the author then goes on to explicate and define two relevant kinds. A virtue reliabilist (Riggs uses the

term "truth conducivist") might define the justification condition on knowledge in such a way as to preclude the kind of chance that Riggs simply calls *epistemic luck*: "*S*'s coming to hold a true belief, *p*, is a matter of *epistemic luck* for *S* to the extent that *p* is unlikely, given that it was produced by process *R*."[24] I will use Riggs's terminology in this section, except that I will refer to this first kind of chance as *coincidence*, allowing us to preserve a broader and more conventional use for the term "epistemic luck" in the sections that follow. Riggs's definition of *coincidence* reflects the objective link to truth that is characteristic of the truth-conducivist account of justification. The motivation behind reliabilism centrally involves disallowing beliefs that are merely "lucky guesses" or coincidental truths from counting as knowledge. Consideration of Gettier cases, for instance, has strongly pushed toward external constraints on knowledge that can prevent a justified, true belief from counting as an instance of knowledge if its truth is merely a coincidence and is not "linked" with its causal ground in an appropriate (epistemizing) way.

By contrast, the kind of chance that responsibilists want to preclude from counting as knowledge is what Riggs refers to as "epistemic accident."[25] Citing John Greco and Hilary Kornblith as examples, he takes their accounts of justification as primarily concerned with the evaluation of epistemic agents or their practices. The responsibilist begins from the agent's motives, and so insists that a belief is justified only where the *intentions to have true beliefs and to avoid false ones* have played appropriate grounding roles in the agent's belief-forming practices. What the responsibilist means by an appropriate grounding role Riggs takes to be a certain kind of *causal* role, and so "*S*'s coming to hold a true belief, *p*, is an *epistemic accident* for *S* if coming to hold *p* was not (sufficiently) caused in an appropriate way by *S*'s intention to have true beliefs."[26]

The responsibilists' motivation for placing such a constraint precluding epistemic "accidents" from counting as knowledge is partly explained by their intuition that the justification condition demands beliefs held (or propositions accepted) not only in conformance with our epistemic norms, but *from* or *in light of* such norms.[27] An agent's consciously held belief that satisfies this constraint will be one that is *guided by* his taking a support relation to hold between reason state *r* and his acceptance of *p*. This concern to preclude *epistemic accidents,* as Riggs points out, also explains the enthusiasm that responsibilists have shown for cases of deviant causal chains, where a reliable BCP has been in operation, yet a belief produced by it fails to "link up" with the agent's intentions or subjective justification in an appropriate (epistemizing) way. Summarizing Riggs's analysis briefly, *coincidence* (or epistemic luck for Riggs) is primarily a matter of success despite low-likelihood, given the agent's actual process, while *accident* is primarily a matter of success despite intentional inefficacy, given the agent's intentions.

Riggs's most challenging claim in assessing the upshot of his analysis is that "The responsibilist and truth-conducivist conceptions of justification define distinct epistemic evaluations and so are not in any interesting sense rival notions."[28] He views their divergent constraints on chance as mutually consistent, deriving from "an identical epistemological concern." From this view of their consistency, he proposes an intriguing conjunctive formula for the justification condition on knowledge. The

necessary and sufficient conditions for knowledge are according to this formula true belief, and a revised third (justification) condition, where the conjunction of the former two conditions "is not a matter of chance" in *either* of the two senses we have specified. This conjunctive definition is interesting in its own right and may prove more resistant to the usual roundup of counter-example cases than definitions that do not include *both* the reliabilist and responsibilist constraints. He proposes a new focus for epistemology in line with his approach, that of working out "an adequate account of epistemically responsible belief, as well as an account of sufficiently truth-conducive belief." As we have previously seen, such a twofold approach is also indicative of the mixed account of justifications widely advocated among virtue epistemologists.

The compatibility that Riggs asserts to obtain between reliabilism and responsibilism is quite appealing; it must be correct in some sense for a mixed account of justification to be plausible. The motivation for his conjunctive strategy also seems clear enough: reliabilist justification prevents a true belief that will constitute knowledge from being a *coincidence*, but not from being an *accident;* and responsibilist justification prevents a true belief that will constitute knowledge from being an *accident*, but not from being a *coincidence.* However, at the same time I have reservations about Riggs's conjunctive strategy for defining knowledge—that is, the simple adding of the one constraint to the other in order to render the account of justification more adequate or 'complete.' This strategy may be shown to be problematic if it invokes assumptions that make it more rather than less difficult to respond to the challenge of skepticism. The one assumption I would focus on here is that of the incompatibility of luck and knowledge. We should notice, after all, that the analysis he presents puts us nowhere closer to explaining the commonplace belief with which we began, that the reliabilist takes herself to be *more amenable* to epistemic luck than the responsibilist. This point should help us to see that using stipulative definitions excluding certain forms of chance cannot provide a suitable answer to the entire complex of problems associated with epistemic luck. As discussions in ethics have rightly pointed out, luck is not entirely a conceptual matter, but depends upon the make-up of ourselves, our world, and our relationship to it. If we all rely on luck in our intellectual lives, then we cannot merely stipulate that knowledge excludes it—not, at least, without simply begging the question against the skeptic by assuming from the outset that we know what we think we know.

I do not mean to burden Riggs with holding the incompatibility thesis in its most objectionable form, the claim that knowledge is incompatible with luck *simpliciter;* he has defined the specific senses to be excluded, and it may be only a limitation of his focus that he does not address other senses of luck which may be compatible with knowledge. But an adequate account must go beyond showing that excluding epistemic coincidences is compatible with excluding epistemic accidents, to a *positively* stated account both of this compatibility and of epistemic justification. This would be one that explains in positive terms what it means to be "in a position to know" and what it means to be a "well-motivated and responsible" agent. Riggs hints at such a positive account in saying that the constraints against the different kinds of chance "arise from an identical epistemological concern," and that the parties to the

debate share a common conception of what epistemic justification is *for* (that is, a kind of *filter* of knowledge claims). Related claims that Riggs makes—that the divergent constraints on justification represent *value-charged demands for the intelligibility of knowledge*, and that neither can claim logical priority over the other—also hint at arguments for their compatibility. Yet these comments far outstep the arguments Riggs actually provides in his article.

Such limitations to his analysis are not surprising, I want to say, because there is another whole side to the issue of luck we have not yet explored. We begin to engage it in the next section by directing critical attention to the thesis of the incompatibility of knowledge with luck. There I will briefly lay out an argument against the incompatibility thesis, showing its self-defeating nature. Then in Section IV, I will further elaborate certain senses of "luck" to which the reliabilist and the responsibilist must each respectively acquiesce once the incompatibility thesis is abandoned. In acknowledging kinds of luck that cannot be excluded but must instead be part of any adequate account of knowledge, I hope to build upon the idea of constraints on justification as value-charged demands for the intelligibility of knowledge, and thereby to clarify the basis for a positive account of the compatibility of reliabilism with responsibilism.

III. Skepticism and the Incompatibility Thesis

In this section I make use of another thought-provoking essay on epistemic luck, "Is Luck Compatible with Knowledge?" in which Mylan Engel argues that "both internalist and externalist epistemologies lead directly to skepticism when they are coupled with the incompatibility thesis."[29] Why must both internalist and externalist epistemologies reject the incompatibility thesis? Because the view that luck *simpliciter* is incompatible with knowledge entails skepticism. But why in turn is this the case? Because each of these theories can in fact be shown to *entail* a kind of luck. Luck, as Nagel pointed out for us, has always provided the skeptic with a foothold; unrealistically denying luck's role in knowledge may provide him a ladder! For the sake of clarity I will distinguish the kinds of epistemic "chance" asserted in the previous section to be incompatible with knowledge from forms of epistemic "luck" whose nature we have yet to explore.

Engel's essay utilizes the distinction between internalist and externalist epistemologies rather than the more subtle distinction between responsibilist and reliabilist epistemologies, which is our real concern. We will need to follow Engel in this for the present, and in order to underline differences, I will typically mean by the term "externalism" its more "pure" or non-mixed versions unless I specifically refer to a "mixed" version. We can also pick up the metaphors with which we began, by examining these kinds of luck in terms of the ideas of perceived crucial discontinuities or inferential lacuna ("gaps"), and responding to these, attempt to build theoretical connections ("bridges").

Gaps are well-acknowledged in ethics, a prominent example being the gap between being virtuous and living well, according to Aristotle: Uncontrolled happen-

ing (moral luck) can step into this gap, impeding the person of *arete* from finding its proper fulfillment in *eudaimonia*. Now it is clear that for internalist accounts in epistemology, there must remain a gap between justified belief and truth. Internalists maintain that epistemic justifiedness is *exclusively* a function of the cognizer's internal states—belief, memory, or perceptual states. No internalist theory can provide a conceptual bridge or connection between justification and truth, as Gettier-cases purport to show. According to Engel then, "no internalist theory can eliminate the role luck plays in a person's coming to have a true belief. So if luck really is incompatible with knowledge, then no internalist epistemology can give rise to knowledge."[30] Since the experience of those in a demon world is subjectively indistinguishable from our own, the internalist is content to construe demon world inhabitants justified in their beliefs if the quality of their efforts to attain truth and avoid error have been beyond impeachment. The internalist takes beliefs as virtuously held on the basis of evidence available to a situated agent, and when it is revealed that those beliefs are nevertheless false due to the demon's systematic deception, she can merely acquiesce to a form of epistemic luck, conceding that luck must intervene to turn our justified beliefs into true beliefs, and hence into knowledge.

Externalist theories may appear to be on much stronger ground, but as Engel points out, if this is so, the reason cannot be that the externalists need not admit luck into their accounts. On the contrary, "externalist epistemologies with truth-connected theories of justification simply replace one kind of epistemic luck with another, for while it is *not* a matter of luck when a [process-reliabilist]-justified belief turns out to be true, it *is* a matter of luck when a belief turns out to be [process-reliabilist]-justified."[31] Engel terms the kind of luck to which the internalist must acquiesce *veritic luck*, and that to which the externalist must acquiesce *evidential luck*. Veritic luck might most simply be understood as luck with respect to the *output* of our BCPs and *evidential luck* as luck with respect to the empirical or evidential *input* that our BCPs must work from and with.

A demon-world case can also be used to illustrate evidential luck. Consider yourself in relationship to your *doppelganger* in the demon world and reflect on the fact that the two of you live phenomenologically indistinguishable cognitive lives. Since there is no discernible difference between your worlds, you must conclude that it could have just as easily been you who were the one with the unreliable belief-forming cognitive processes. Given your situated place within this scenario, there is no evidential basis that could possibly serve you as a guide from which to reach any other conclusion. If this is so, then it is still a matter of luck that, in the context of this scenario, it is you and not she who has the process-reliabilist *true* beliefs.

According to the foregoing argument, the externalist must acknowledge that it is always a matter of *evidential luck* when reliably produced beliefs turn out to meet the subjective conditions that internalists would place upon knowledge. And the internalist must acknowledge that it is always a matter of *veritic luck* when our beliefs turn out to meet the objective conditions an externalist would place upon knowledge. But is this a cogent line of argumentation? Clearly Engel has allowed the very basis for the characterization of luck to differ in his demon-world scenarios. But

why should he not? Perspective matters! If it takes an externalist perspective to point out the kind of luck which internalism entails, it seems valid to allow the same in reverse, by allowing the evidential factors rather than matters of objective fact to be the ones that may vary in the examples. The perspective focusing on the evil-demon's revelation to me illustrates that it must always be a matter of luck *from an external or factual perspective* when an internally justified belief turns out to be true. The perspective focusing upon my relationship with my doppelganger illustrates that it must always be a matter of luck *from an internal or evidential perspective* that one of us, myself or my double, should in fact be the one possessing sound evidence for normally functioning BCPs to use as input.

But are these kinds of luck of equally significant import for epistemology? In Gettier and other types of cases which might have been discussed in connection to the kinds of "chance" examined in the previous section, our intuitions tell us that an agent's justification is undermined by certain facts about his situation that are beyond her ken. In the eyes of their devisers, at any rate, such examples exploit *factual defeaters* to an agent's knowledge claims that, once they have done their dirty work, leave the would-be knower only with what Matthias Steup calls *nonepistemizing justification.* But is that also the case with the kinds of demon-world scenarios we are now considering? That seems not as clear. We are now encountering forms of epistemic luck that have more relevance to the general challenge of philosophical skepticism than to the statement of specific conditions on knowledge. This is partly why I separate them here from Riggs's two forms of *chance* discussed in the previous section.

Our two forms of luck both have important epistemic implications, even if they differ from the forms of chance we earlier investigated. Engel himself argues that only *veritic luck* has epistemological import, thereby giving the strong advantage to externalism. I disagree here: *Evidential luck*, too, is a kind of luck that concerns the crucial relationship between the agent and the known fact; it is not, any more than is *veritic luck,* a kind that concerns merely the existence of the known fact or the existence or abilities of the man who knows. I will need to say more on this later, but I take input and output luck to be on generally equal footing in this regard. It takes the other's perspective to *point out* the type of epistemic luck that each account must admit, and internalists and externalists both think that once pointed out, the kind of luck systematically implied in their adversary's account represents a deep *fault* in their approach. Moreover, both perspectives contribute to explicating the pretheoretical notion of luck as referring to factors affecting that crucial relationship (between the agent and the known fact) that remain *beyond our ken* or, in other words, beyond human control.

Since internalism and externalism are defined by Engel in the usual mutually exclusive and exhaustive way (exhaustive except for the option that would radically eliminate justification as a condition on knowledge altogether), we are left with the implication that luck is deeply involved in knowledge on either account. So to return to our main line of argument, if luck *simpliciter* really were incompatible with knowledge, then *both* theories would make knowledge impossible and hence lead directly to skepticism. This argument I believe is cogent, and it leaves us with the

following options: (1) embrace skepticism, (2) embrace a non-justificationist account of knowledge, or (3) reject the incompatibility thesis.

If we accept this scenario, then the last option appears most promising. This means that, again following Engel's lead, we must try to "reconcile the rather strong intuition that epistemic luck [chance] is not compatible with knowledge with the equally evident observation that it must be."[32] Both accounts have been only partially successful in shielding knowledge from a kind of luck that its adversaries see as undermining its claim to adequacy. Of course, there are certainly kinds of luck that are simply irrelevant to the question of knowledge. But neither of the types of luck Engel discusses appears to be of such a type, and the route of a non-justification form of externalism is unappealing. Nor should we assume that the presence in any degree of one or the other kind of epistemic luck is sufficient to undermine knowledge, an assumption which might lead us to embrace skepticism. Consistent with Engel's third option of rejecting the incompatibility thesis, the point we should stress is rather that on both internalism and externalism, understood as mutually exclusive accounts, it is necessary for luck to step into a perceived "gap" in order for justified belief and truth to link up in an epistemizing way. Both theories, in this sense, remain essentially "incomplete" in their response to skepticism.

We have now argued that the incompatibility thesis is mistaken. A strictly conjunctive account, we should now be able to see, is unrealistic as an attempt at bridge building. It requires too much of human cognizers, because it excludes from the analysis the influence of factors that are a matter of luck. An adequate account of epistemizing justification cannot be stated in Riggs's negative fashion of merely excluding luck from knowledge, even where that statement is in fact a conjunctive one expressing *both* internalist and externalist chance-precluding constraints. We have also argued that pure internalism and externalism—theories that state their conditions for knowledge in mutually exclusive terms, must each by their own respective logic "acquiesce" to an important role for epistemic luck. Since each is premised upon the incompatibility of luck with knowledge, this acknowledgment paralyzes their ability to state conditions of knowledge in a positive manner. The upshot (here in line with Riggs) is that internalism and externalism as so understood are necessarily incomplete accounts, and that neither alone can be an adequate response to the skeptic. Taking this as background, we are now in a position to discuss the advantages of the *mixed* externalist approach to justification and perhaps outline the demands upon the *positive* account of epistemic justification we are seeking.

IV. The Gaps Problem: Dealing with the Incompleteness of Epistemic Theories

It is in their acquiescence to different forms of luck that I want to say we find the deepest source of the divergence between externalists and internalists, and by extension, between reliabilists and responsibilists. In the context of our earlier discussion, we said that gaps represent perceived inferential lacunae or discontinuities in the human cognitive process. Gaps are present to us existentially simply as forms of luck

that demand our recognition because they impact our lives. Philosophically, and for better or worse, we tend to perceive these demands differently one to another. If it weren't obvious to the reader already, the discussion of the previous section serves to show how epistemologists routinely view the forms of luck logically implied by their own favored position as epistemologically innocuous, and their adversary's forms as nothing short of a plague upon their house. This disemblance cannot long be maintained, and the oft-noted "stalemate" between these mutually exclusive adversaries is easy to recognize. By admitting any gap, any discontinuity in the theoretical account of human cognition, internalist and externalist epistemologies reveal themselves as incomplete. Since the forms of luck that the internalist and externalist each must acquiesce to are the direct target of their adversaries, it becomes obvious to those who are party to the debate that neither approach, freestanding, is adequate to respond to, let alone capture both sets of intuitions.

Conceptualizing this incompleteness in terms of "gaps" is simply meant to formalize this problem. Gaps now *correspond* to acknowledged forms of luck impacting human cognition and potentially blocking epistemizing justification. Bridges are now theoretical attempts, either conceptual or empirical, to *respond* to gaps through our philosophical reconstructions, and to successfully carry through the claim that present justification is indeed epistemizing. To clarify this, we must first try to formalize our conception of the gaps themselves. Internalist theories, we've shown, leave us with a gap between (1) truth and (2) justified belief. We can simply call this 1–2 gap the *veritic gap* to correlate with the form of luck to which internalists must acquiesce. Externalist theories, however, do provide a conceptual link between justified belief and truth; that is to say, a conception of a necessary connection between knowledge and its object. But for them a gap re-opens between belief (or more strictly, propositional acceptance) and (3) good reasons or adequate evidence. This is a token of a discontinuity between the reliabilist's strong *conceptual* claims and their lack of a supporting theory of evidence.[33] Borrowing a Greek term to balance against the Latin *veritas,* we can call this 2–3 gap the *zetetic gap.* This comes from *zetesis,* referring to the inquiring *quality* of an investigator, or of a method of investigation, and was used by the Greeks in the context of a common search for an *unknown* truth. Again this is intended to match the sense of *evidential luck* discussed earlier.

Our characterization of gaps follows the analysis of our previous section. But we should add at least one more, that between (3) reasons and (4) (proper) motivation. We can call this 3–4 gap the *enkratic gap*, for the Greek term for continence. Do reasons have to motivate? For instance, if I accept that I have reason to doubt/believe some proposition, must I be *motivated* to some degree to doubt/believe that proposition? What Christopher Hookway calls (motivational) internalism answers "yes"; he sees the normative impact of reasons as inseparable from our motivational tendencies, at least in rational agents. For the internalist, epistemic rationality requires us to be rightly motivated: Evidence doesn't provide me with reasons, unless it is integrated in an appropriate way with my motivational tendencies, and "it seems that the existence of a gap between my normative judgments and my motivations signals a kind of irrationality."[34]

Now if we hold this model of three gaps between four terms [1–2–3–4] in mind, we can say that the amalgam of views associated with internalism allows it to present us with bridges for *both* the zetetic and enkratic gaps, but at the cost of leaving the veritic gap *radically open*. Conversely, it is often said that externalism presents a great departure from "traditional epistemology" with its predominantly internalist character. In the terms we have used in this essay, we might then say that externalism has the advantage of being able to *radically* bridge the veritic gap, but at the cost of leaving us with suspect resources for addressing the other two.

To conclude our brief discussion in this section, I would point out that each of the gaps and associated kinds of epistemic luck we have discussed have strong analogues in Nagel's discussion of moral luck. The veritic gap (and with it Engel's *veritic luck*) bears resemblance to what Nagel termed *consequential luck*; the zetetic gap (and with it Engel's *evidential luck*) can be profitably compared with Nagel's *circumstantial luck;* and finally, the enkratic gap, as exemplified in the possibility of a kind of epistemic *akrasia* that Hookway examines, betokens largely unexplored comparisons with Nagel's *constitutive luck*. What I will call the *gaps problem* is the problem of providing a *positive* account of knowledge and justification, one that addresses each of the discontinuities under a single unified perspective, without ignoring or negating the influence of factors that are a matter of luck. Thus the parallels suggested between acknowledged forms of moral luck and the forms of epistemic luck we have examined are further indication of the import of the *gaps problem* for epistemology today.

V. Mutual Assistance

We have now described two specific forms of "chance" (section II) that, though sharply contested among reliabilists and responsibilists, are each claimed by some to preclude true beliefs from constituting knowledge; we have also described forms of "luck" (Section III) that must be acknowledged rather than precluded and that, while also impacting epistemology, do so primarily in terms of their relevancy to a general skeptical challenge that virtue epistemologists, whether those we have called virtue reliabilists or responsibilists, must confront together. In section IV we have seen that the presence of veritic and evidential luck indicates an incompleteness in internalist and externalist epistemologies and, most important, we have gone on to connect each form of luck with a theoretical "gap" whose place in any positive account of justification demands acknowledgment. In this final section we will discuss objections to the present analysis of epistemic luck and clarify the demands on a mixed externalist account of justification in addressing what we have come to call the *gaps problem*.

One objection that might be brought against the discussion of epistemic luck in the previous two sections is simply that it seems to bear little direct implication for virtue epistemology. After all, the arguments in those sections lean heavily upon the dichotomy between internalism and externalism, understood as mutually exclusive and exhaustive accounts of justification. But we had previously been at pains to

show how and why virtue epistemologists have been attracted to a "mixed" externalist account, one that is capable of consistently integrating responsibility constraints on justification that ensure the guiding role of good reasons in reflective human knowledge. To hold a mixed account is already to agree that purely internal or purely external constraints cannot provide the right conditions for *epistemizing* justification. Even if there are serious differences among those who advocate mixed accounts, those we began by characterizing as virtue responsibilists are not really internalists, since they certainly do not think that the conditions on justification are *exclusively* internal; and those we characterized as virtue reliabilists do not claim, as some externalists might, that we can do without an account of subjective justification for reflective knowledge. So the discussion became skewed when we slipped back into talking in terms of a dichotomy between internalism and externalism that most of us agree has outlived its usefulness. For one who accepts a more compromising mixed account, the problem of epistemic luck need not have either the decisive importance or the divisive character I have suggested it has.

In response, I would first have the reader note that I have been concerned with an *indirect* effect of the problem of epistemic luck on mixed externalist accounts. We have considered how the understanding of epistemic luck in virtue epistemology tends still to be divided along lines that reflect a backdrop of externalist/internalist debate in epistemology over the past three decades. I agree that this distinction is becoming outmoded and hope that the distinction between reliabilist and responsibilist virtue epistemology will before long also outlive its usefulness. But that should not preclude our notice of real and present tensions among those working in the field.

Second, two common claims made by virtue epistemologists of all stripes are (1) that their approach has something unique to offer in the way of overcoming the opposition between internalist and externalist conceptions of justification, and (2) that it provides substantial resources for addressing the challenge of skepticism. I can find no better way to put these two common claims to the test than to inquire into the reception and understanding of epistemic luck. A virtue-based account of a mixed character should show that justification does not turn on the issue of "internal accessibility" to the ground of one's belief, the issue that the distinction between internalism and externalism was build around. While the various strategies of virtue epistemologists who advocate mixed accounts of justification suggest a displacement of the access issue, we have found that the issues have been transformed but not resolved, and that serious disagreements persist between virtue reliabilists and virtue responsibilists.

Third and most important, I disagree with Driver when she says that mixed accounts represent inherently unstable compromises and that "the superficial plausibility of the mixed account is purchased at the cost of significant theoretical advantages" presented by the "pure extremes" of internalism and externalism. On the contrary, I have tried to argue that the advantages reside in a view that takes seriously both reliability and responsibility constraints on justification, because this mitigates the necessary incompleteness of 'pure' internalism and externalism and provides that best basis for responding to skepticism.

But I would make a partial concession to Driver's complaint: Mixed accounts certainly are not immune to the *gaps problem* or to more generalized problems concerning epistemic luck; they remain subject, as Driver claims, to many of the *same* theoretical problems that afflict the internalist and externalist extremes. This concession may indeed be implied by my view that one simply can't bridge all three of our recognized gaps simultaneously under one perspective (as the notion of adding or conjoining responsibilist to reliabilist conditions on justification would suggest). This is philosophically unrealistic because it would produce a conception of the human agent with cognitive powers far in abundance of what the cognitive sciences and our own lived experience reveal to us. To say as we did before, that Nagel's three kinds of ethical luck can be seen to have important epistemic analogues, is to say that a mixed account is one that will seek to understand the epistemic import of *each* of the three gaps we have sketched and to produce a positive account of epistemic justification *in light of them*.

Acknowledgment of the *gaps problem* motivates rather than prevents the tasks of theoretical bridging. But my conception is one on which there are serious trade-offs involved in such theoretical endeavors: to emphasize a conceptual connection here is to allow a potentially larger lacuna elsewhere in one's system of thought. And if this is correct, then there are undoubtedly a variety of different ways that epistemic luck can be theoretically addressed. One might say that our theoretical bridges are makeshift, especially in this era of new empirical studies in the cognitive sciences, and that they will need to be constantly adjusted one to another as we learn more about human cognition and motivation structures and as we advance our theories of knowledge and evidence in light of them.

It is simply not clear at present how well mixed accounts can handle this task, but we have shown reason to think that mixed accounts are on better footing than non-mixed accounts in this regard. While sometimes lauded for a certain perceived rigor or inner consistency, we should now be able to see that non-mixed approaches place themselves in a position of having to argue, most implausibly, that the type of luck to which their account must acquiesce lies outside the relationship between the knower and the known and is of no real epistemological significance. To return to Driver's objections, I cannot resist playing her claim of *inherent* instability off against Jonathan Dancy's similar claim. An irony is present because Dancy uses similar reasoning to prescribe a move *away* from a consequentialist/reliabilist conception of the virtues toward a virtue epistemology based upon Aristotle's account of the moral virtues:

> Non-consequentialists also are unwilling to admit that the consequentialists are right about anything, because they feel that consequentialism is like a cancer — once one has let it in at all it will grow until it has taken over completely. The crucial question is whether the two camps are right at least about this, that no compromise is intellectually acceptable. And I think that they are.[35]

Though the conclusions that Driver and Dancy reach are highly antithetical, their 'inexorable' logic on this score is strongly analogous, since both argue that compromise solutions are theoretically unstable and intellectually unacceptable.[36]

Such claims about theoretical instability really amount to little more than *predictions* concerning the fate of mixed accounts. If mixed accounts prove only superficially plausible as both think, virtue reliabilists and responsibilists can be expected to soon part company and to do their research under separate self-descriptions rather than accepting their shared description as virtue epistemologists. With such predictions I need not agree. If this prediction is *not* to prove correct, however, there is a burden on the virtue epistemologist to present a theoretically consistent, positive account of epistemic justification, one that reliabilists and responsibilists both substantially agree with and contribute to. At the end of a lengthy essay, however, it may come as some relief to the reader to learn that I do not have such a theoretical account ready to offer!

I would, however, put forward a practical point in closing. The stability of any research program is often as much a practical matter as a theoretical one. The easiest way to exhibit instability is to lose the valuable benefits of mutual assistance. Reliabilists have focused on the demands of philosophic naturalism and on conceptions of the supervenience of normative properties; they have been especially attuned to the role of the cognitive sciences in understanding both reflective and non-reflective (including animal) knowledge, and to the social and communal dimensions of epistemic evaluation. Responsibilists have focused on concerns with active agency in the context of reflective knowledge and with studies and thick-descriptions of particular epistemic virtues and vices. They have been especially attuned to the interconnections between ethical and epistemological disposition and to the dynamics of individual psychology. With a mixed externalist account of justification progressing as a shared project among virtue epistemologists, I am hopeful that the complementarity of these research foci will achieve greater recognition and that the interests that shape research within virtue epistemology will merge significantly further. Surely this mutual-acknowledgment and integration of interests is practically necessary if virtue epistemologies of a "mixed" externalist character are to confront the very serious doubts about their theoretical stability that we have heard expressed from positions representing internalist and externalist extremes.

Notes

1. Claudia Card*, The Unnatural Lottery: Character and Moral Luck* (Philadelphia: Temple University Press, 1996). See also Daniel Statman (ed.), *Moral Luck* (New York: SUNY Press, 1993).

2. For a discussion of this division in virtue epistemology in light of debate over Aristotle's own distinction between the moral and intellectual virtues, see my "The Role of the Intellectual Virtues in the Reunification of Epistemology," *The Monist,* 81, 3 (1998): 488–508.

3. Julia Driver, "Moral and Epistemic Virtue," in Guy Axtell (ed.), *Knowledge, Belief and Character: Readings in Virtue Epistemology* (Lanham, Md.: Rowman & Littlefield, 2000), 123–34. Compare Driver's description of her view with Firth's term "epistemic rule-utilitarianism," which Sosa has more recently used to describe reliabilist-oriented virtue epistemology.

4. Linda Zagzebski, "Religious Knowledge and the Virtues of the Mind," in *Rational*

Faith: Catholic Responses to Reformed Epistemology, Linda Zagzebski, ed. (Notre Dame, Ind.: University of Notre Dame Press, 1993), 202.

5. See also Ernest Sosa's *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991) and Linda Zagzebski's *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996). By received definitions, internalism insists that the conditions on justification are *exclusively* internal, and externalism denies this, so any mixed account will by definition be classed as externalist.

6. Sosa, *Knowledge in Perspective,* 28.

7. Stanford Online Encyclopedia, entry for "Virtue Epistemology."

8. Lorraine Code, *Epistemic Responsibility* (Hanover, N.H.: University Press of New England, 1987), 51.

9. For more on this distinction, see my "Introduction" to *Knowledge, Belief and Character*, xi–xxix.

10. For another article defending "mixed" accounts of justification that makes some closely related points, see Paul Bloomfield, "Virtue Epistemology and the Epistemology of Virtue," *Philosophy and Phenomenological Research* 60, 1 (2000): 23–43.

11. Montmarquet's thesis of *direct* as opposed to indirect doxastic responsibility disputes the reliabilist assumption that control over beliefs can be achieved only directly, through mental *actions* of one sort or another. Montmarquet rejects the acceptance/belief distinction and argues for (soft) *doxastic voluntarism*. See his *Epistemic Virtue and Doxastic Responsibility* (Lanham, Md.: Rowman & Littlefield, 1993), 90–95. For a pragmatist account supporting the acceptance/belief distinction, see D. S. Clarke, Jr., *Rational Acceptance and Purpose* (Lanham, Md.: Rowman & Littlefield, 1989), especially 33–36.

12. Compare Zagzebski, *Virtues of the Mind*, 59–73.

13. Thomas Nagel, "Moral Luck," in *Mortal Questions*. (Cambridge: Cambridge University Press, 1979), 24–38.

14. Zagzebski, *Virtues of the Mind*, 71.

15. Zagzebski, *Virtues of the Mind*, and Montmarquet, *Epistemic Virtue* are again prime examples of this responsibilist perspective. For more on the issue of defining the intellectual virtues, see my "Virtue Theory and the Fact/Value Problem," in *Knowledge, Belief and Character*, 177–203.

16. Richard Paul may be a holder of the strong thesis: "The problems of education for fair-minded independence of thought, for genuine moral integrity, and for responsible citizenship are not three separate issues but one complex task." Richard Paul, "The Contribution of Philosophy to Thinking," in *Critical Thinking* (Sonoma, Calif.: Foundation for Critical Thinking, 1993), 424–25.

17. Zagzebski, *Virtues of the Mind,* 255: "I think of this move as expansionist rather than reductionist since it would be more accurately described as expanding the range of ordinary moral evaluation to include epistemic evaluation, rather than reducing the latter to the former." Note that Greco describes Zagzebski's account as "'neo-Aristotelian' rather than 'Aristotelian,' because Aristotle did not hold that the moral and intellectual virtues are unified in this way" (Stanford Online Encyclopedia). But Driver's attempt to frame examples where moral goodness is incompatible with epistemic virtue poses no apparent problem for Zagzebski, her explicit target, who explicitly refrains from taking a position on the classic Socratic/Aristotelian unity thesis (*Virtues of the Mind*, 156–157).

18. James Montmarquet, "An 'Internalist' Conception of Epistemic Virtue," in *Knowledge, Belief and Character*, 135–48.

19. Zagzebski, *Virtues of the Mind*, 258; Bloomfield, "Virtue Epistemology," 35.

20. Driver, "Moral and Epistemic Virtue," 123.

21. Bernard Williams, *Moral Luck* (Cambridge: Cambridge University Press, 1981), 26. For a related critique of neo-Kantian elements in Zagzebski's thought, see Amelie Rorty's

"Distinctive Measures of Epistemic Evaluation" and Zagzebski's response in the book symposium on *Virtues of the Mind,* in *Philosophy and Phenomenological Research* 60, 1 (2000): 203–206, 216–19.

22. Nagel, "Moral Luck," 467.

23. Wayne D. Riggs, "What Are the 'Chances' of Being Justified?," *The Monist*, 81, 3: 452–72, 463.

24. Riggs 465. For background see also Peter Unger's "An Analysis of Factual Knowledge," *Journal of Philosophy* 65, 6 (1968): 157–70.

25. Riggs, "What Are the 'Chances,'" 467.

26. Ibid.

27. If I can use Robert Audi's guide/effect distinction to connect Riggs's sense of causal efficacy with *believing for a reason*, it seems that a constraint against accidents is one that should disallow knowledge in cases where an agent *S*'s belief *p* may be merely the *effect* of a reason state *r*, but not also *guided* by reason state *r*. "But if *r* is a reason *in the light of* which *S* believes that *p*, guides *S*'s belief formation (or retention), and is discriminative, then surely the explanatory relation between the belief that *p* and the basis and connecting beliefs is not accidental." Robert Audi, "Belief, Reason, and Inference" in *Philosophical Topics* 14, 1 (1986): 47.

28. Riggs, "What Are the 'Chances,'" 453.

29. Mylan Engel, "Is Epistemic Luck Compatible with Knowledge?," *Southern Journal of Philosophy* 30, 2 (1992): 59–75. For a response to Engel quite different from my own, see Barbara Hall's "On Epistemic Luck," *Southern Journal of Philosophy* 32 (1994): 79–85.

30. Engel, "Is Epistemic Luck Compatible?," 61–62.

31. Ibid., 63.

32. Ibid., 67.

33. Ibid.

34. Christopher Hookway, "Regulating Inquiry: Virtue, Doubt, and Sentiment," in *Knowledge, Belief and Character*, 149–160.

35. Jonathan Dancy, "Supervenience, Virtues and Consequences," reprinted in *Knowledge, Belief and Character*, 73–86.

36. Dancy continues: "Consequentialists are wise to seek to give a unified account of all the virtues, because otherwise they will find themselves saying that of the virtues, some are virtues for one sort of reason and others are virtues for another. This position is theoretically unstable, and will always be vulnerable to one which manages to give the same account of why this or that feature is a virtue throughout. Similarly, virtue theorists are right to resist the irruption of a second form of explanation of the status of a character trait as a virtue. Their standard form, which asks how the virtues *together* contribute to a good epistemic life, is perfectly capable already of capturing the nature and role of the consequence-related virtues. That they are consequence-related does nothing to show that we should accept a consequentialist understanding of them" (82).

11

EPISTEMIC *AKRASIA* AND
EPISTEMIC VIRTUE

Christopher Hookway

Introduction

Ordinary practical *akrasia*, incontinence, or weakness of the will occurs when some-
one consciously or deliberately makes a choice that she sincerely believes is wrong.
She decides to continue smoking in spite of sincerely acknowledging that it would
be best, all things considered, to give up; or she reaches for a cigarette while fully
aware of her resolve to smoke no more. Such phenomena are a distressingly com-
mon feature of most lives, yet philosophical theories of practical rationality, action,
and evaluation can have considerable trouble escaping from the conclusion that they
are simply impossible. This makes study of practical *akrasia* an especially useful
technique for uncovering important and heretofore unaccounted-for complexities
in the structure of practical reasoning.

   This essay is concerned with whether a parallel situation arises in the study of
theoretical rationality. Are there also cases of epistemic, doxastic or theoretical *akra-
sia*? In the most full-blooded form, such cases would occur if someone consciously
accepts some proposition while also accepting that it is epistemically wrong to do
so—perhaps she thinks that there is strong reason to accept its negation. This essay
discusses why such phenomena can seem so problematic (section 4) but argues that,
once we look closely at the structure of theoretical reasoning and inquiry, we can
make sense of forms of irrational belief that are closely analogous to akratic action
(section 5), and, indeed, there are many philosophically interesting examples of this
(section 6).

   Although these issues are of intrinsic interest, they are examined here for the
sake of the light they cast on some general issues about epistemic evaluation. In-
quiries and deliberations are activities with distinctively epistemic goals: they are di-
rected at solving problems of fact, at finding things out. Our epistemic normative
standards are reflected in the ways in which we carry out these activities: Whether
our inquiries and deliberations take us to the truth will depend in part on how skill-
fully we control their progress and on the adequacy of the standards we employ in

doing so. Making sense of epistemic *akrasia* will require us to understand some important complexities in the structure of inquiries and deliberations and to see what sorts of capacities we must possess if we are to take responsibility for how well they are carried out. I shall conclude that if epistemic *akrasia* is, indeed, possible, then the normative regulation of deliberation and inquiry requires the agent to possess states or traits that are best seen as virtues. They are states, probably states of character, which are manifested in the ways in which we organize and regulate our activities and which enable us to carry out those activities well or rationally. I hope to argue from the possibility of epistemic *akrasia* to the need to take epistemic virtues seriously in our theory of epistemic evaluation.[1]

Section 2 explores some cases of ordinary practical *akrasia*, pointing out some features of practical deliberation and inquiry and emphasizing the most important features which should carry over into plausible examples of epistemic *akrasia*. Section 3 then introduces a first weak form of epistemic *akrasia*: If activities can be carried out akratically and inquiries (and deliberations) are activities then, it seems, the latter can also be carried out akratically. This provides background for the ensuing discussion of full-blooded epistemic *akrasia*.

Ordinary *Akrasia*: Some Examples

Philosophers' characterizations of *akrasia*, of ordinary incontinence, are surprisingly varied. Some describe the akratic as performing one action while acknowledging that there is better reason to perform another; others as performing an action while acknowledging that the reason for doing so is inadequate. Some identify the focus of *akrasia* as *choice* against one's own best judgment (Wiggins 1987: 240), others as *action* (de Sousa 1987: 199, Rorty 1983: 175, Davidson 1980: 21, 22), and yet others as *intention* (Williams 1990: 120). These different formulations need not be inconsistent: if all are failures of "continence," if all exhibit lack of the same virtue, then they can be taken as different examples of a related set of phenomena. All involve a failure of sincere value judgment or commitment to have an appropriate influence upon the processes of deliberation and action. And the challenge they raise concerns how something can indeed *be* a sincere value judgment of mine if it is not manifested in my deliberations and actions.

A preliminary sorting of some of these phenomena can be obtained by taking seriously the fine detail of practical deliberation. Consider an example:

> Imagine someone who believes that it would be good to contribute to alleviating the suffering of famine victims in Ethiopa. On reflection he decides that it would be right, all things considered, were he to do so. The story could then be developed in three ways. Having formed a general resolution to help, his attempts to formulate a more specific intention—perhaps to write a large check to Oxfam or to some other agency—all somehow fail. The general resolve never turns into a more specific intention. Or having decided to contribute to Oxfam, and having decided upon the amount, he finds that he is always diverted by other concerns and the check never gets written. Or perhaps the need to pay for his holiday gains

urgency and he readily (too readily) decides to postpone his donation for yet an-
other month. And in all these cases, recognition of what he has done can be a
cause for embarrassment and surprise. Somehow an evaluation he takes to be
genuine, and an intention he takes to be real and sincere, fail to translate them-
selves into action. [Example borrowed and adapted from Velleman (1989: 138.)]

Notice some stages in what is described here:

(1) The agent attaches value to the relief of suffering in Ethopia (*G1*); he judges
that action to this end would be good.
(2) He decides that it would be right or best, all things considered, for him to act
in order to contribute to *G1*.
(3) He forms a general resolution so to act.
(4) He judges that the best means for achieving his goal, all things considered
(or a fully satisfactory means for doing this) is to write a large check to
Oxfam.
(5) He forms a specific intention to write a large check to Oxfam (*G2*).
(6) He retains that intention in the face of irrelevant competing claims upon his
resources.
(7) He acts on that intention.

The three upshots we described above all involved this process being cut off before
the final stage was reached. In the first case (there were two variants) it is cut off after
stage (3) or stage (4), before a more specific intention has been formed. In the other
two cases, the more specific intention was formed but it failed to issue in action: in
the second case, because it was abandoned for evidently poor reasons; and in the
third case, through what is more straightforwardly a failure of will. These different
possibilities (and they may not be exhaustive) all seem to be forms of *akrasia*. The
agent fails to act and deliberate as is required by sincere evaluative commitments.
    The failings fall into two very broad classes. Let us take it that when I make a de-
cision or form an intention, or indeed when I possess any intention at all, I acquire
a distinctive commitment. Unlike some other goals or desires, a commitment is not
just an end that will be weighed in the balance with others when the need for action
arises. Unless a commitment is actually *abandoned* (often for good reasons), it pos-
sesses a kind of authority which prevents our treating it as simply one among a set of
possible ends. The first class of evaluative failings concerns the relations between my
evaluations — including evaluations all things considered — and my commitments.
I may fail to decide or intend to do what I judge it would be best to do; or I may de-
cide or intend to do what I judge that I should not do. If the process described ear-
lier were to halt at stage (2) or at stage (4), it would exemplify this pattern. The sec-
ond class of failings concerns the fate of my commitments. They may just fade away
for no good reason, or I may allow myself to abandon them for what I know to be
bad reasons. Rationality requires me to ensure that my commitments respect my
evaluations. It also requires me to be true to my commitments: abandoning them
when there is good reason to do so; but sticking by them when there is no good rea-
son to abandon them. The other cases all fit this second pattern. According to David

Wiggins, the second form of these provides the prime focus of continence: continence is an executive virtue that enables us to sustain our commitments. And Richard Holton has explained the ordinary notion of weakness of will as fitting the second pattern too (1999). When we turn to cases of epistemic *akrasia*, we must ask whether there are cases falling into each of these classes.

Akrasia, then, involves conflicts between our evaluations and our commitments, or between our commitments and acts falling under them. The most interesting and problematic cases occur when both of the conflicting elements are fully conscious or readily available to consciousness. When we behave akratically, we are often aware that this is what we are doing; this is not a form of behaviour that has, somehow, to be kept secret even from ourselves. The incontinent smoker is fully conscious of her resolve to give up her habit just as she reaches out for another cigarette. Indeed, as she acts, she may be fully confident that she will subsequently feel regret, shame, or even guilt about what she is doing. In this it differs from self-deception: when the jealous man self-deceptively believes that his partner is being unfaithful to him, he cannot be, at the same time, aware that he is utterly unreasonable so to believe. Epistemic *akrasia* should inherit this openness to consciousness from its practical analogue: A simple and extreme case would be one where someone consciously accepts a proposition while "judging" that he is wrong to do so, or even that he has a much stronger reason for believing its negation. For reasons we shall discuss below, the existence of epistemic *akrasia* is much less evident than is the existence of ordinary practical cases. Perhaps it always involves a degree of self-deception—there may be a continuum of cases differing in the degree to which the cognitive operations in question are "open to view."

Why do these sorts of phenomena seem problematic? In the case of self-deception, the difficulties typically stem from the fact that the "self" is involved both as deceiver and as person deceived. This seems to require a mass of beliefs, desires, and projects which must be insulated from each other but sufficiently integrated to count as states of the same person. The challenge is to find a way of thinking about this distinctive kind of evaluation. The problems presented by *akrasia* are different and depend crucially upon the fact that some of the elements involved are *evaluations* and *commitments*. They seem to be cases where I value A more highly than B, yet my choices and preferences, in situations where both A and B are possible appear to betray a preference for B. The relative strength of the evaluations is not reflected in the strengths of my motivations when it becomes time to act. Many views of evaluative beliefs tie the relative strengths of my evaluations to the relative strengths of my desires, of my motivational states. *Akrasia* makes that problematic. Turning to the epistemic case: We need to understand how I can be *sincere* in my judgment that there is insufficient reason to believe p, indeed much better reason to believe *not-p*, when, faced with the need to form an opinion, I acquire the belief that p. Many philosophers assume that accepting reasons involves making evaluations, and they also hold that evaluations are typically reflected in the ways in which we act and form beliefs. What can the relative strength of sincere evaluations be if it is not reflected in the choices that I make?

Akrasia appears to raise issues about motivation. Generally we are motivated to

act in accordance with our evaluations and to conform to our commitments unless we acquire good reason not to do so. The akratic appears to lack this motivation. Issues are thus raised about how this motivation works in the cases where we do act as, we know, we ought; and about how that motivation can be lacking in other cases. Cases of *akrasia* appear to challenge the suggestion that merely understanding and accepting a proposition that expresses a commitment is sufficient to motivate me to act. One merit of appeals to virtue is that they promise to explain how we are motivated to act on our evaluations and commitments.

So when we turn to epistemic cases, we must distinguish phenomena that concern ineffective evaluations from those which concern wavering commitments; we must accommodate the possibility that the conflicting elements can somehow both be consciously present and effective at the same time or within the same process of deliberation; and we must make sense of how the ineffective evaluations and commitments can be genuine and their avowals can be sincere. We shall also expect to face some problems about "epistemic" motivation.

3. Akratic Inquiry

In this section, we consider some less problematic forms of *akrasia* whose content is clearly epistemic but which are distinct from the full-blooded cases we described in section 1. Deliberation and inquiry are themselves activities: We raise questions, make observations, conduct experiments, check proofs, consult colleagues, rehearse arguments to check for fallacies, and so on. These activities are straightforwardly intentional. Whether my beliefs are justified will depend on how well I carry out these activities; my view of how strongly the evidence supports some proposition will itself depend on how carefully I have checked, double checked, consulted other people, and so on. My reasons for collecting new evidence will be practical reasons, reasons for carrying out a distinctive activity. That the goal of the activity is an epistemic one does not undermine this fact.

If inquiries (and deliberations) are activities, then, like other activities, they can be carried out akratically. In that case, it is unproblematic that belief can be akratic: It can be produced or sustained by inquiry or deliberation that is akratic in the ordinary practical sense. I know it is best to make careful checks before accepting scurrilous gossip about a friend; but it does not follow that I will always do so. Aware that my intuitive probability judgments, like everyone else's, are often extremely unreliable, I may formally decide never to trust them. However, when I respond to a run of reds by betting a large sum of money on black, I may persuade myself that *this* judgment is so obvious that the check is unnecessary. And, in doing this, I may be aware that I am failing to conform to important epistemic commitments. Although these are examples of akratically formed belief, they need not involve full-blooded epistemic *akrasia*: Incontinence may prevent my even forming the conflicting judgments that full-blooded *akrasia* would require.

Epistemic *akrasia* can display both of the forms described in the last section: My commitments can fail to conform to my evaluations; and my commitments can fail

to be reflected in how I conduct inquiries and deliberations. I can judge that the available evidence is insufficient to support some belief I hold, or believe that the methods used to acquire it were unreliable, yet still fail to form a resolution to examine the matter further. Perhaps I do not decide to phone the airline to check that my plane is not delayed despite my being aware that delays are common and there would be serious practical implications were I to miss a later connection. And, as in the gambling example of the last paragraph, I can fail to conform to my firm epistemic commitments in particular cases. In each case, it seems no harder to envisage each of the conflicting elements being open to consciousness than they are in prototypical practical cases.

It is easy to see how such lapses could be explained: Wishful thinking may be involved, or a desire to reduce the anxiety of having to face the tasks involved in adjusting my travel plans.[2] So long as wishful thinking involves failing to conform to epistemic standards we take ourselves to be committed to, it may indeed be a form of epistemic incontinence. That may be the case here: we have failed to take account of all the evidence that was accessible or available. And this is explained by reference to the fact we end up without beliefs that we would prefer not to have — we would prefer there to be no delay and thus we prefer not to have to deal with the possibility. But this differs from the full-blooded epistemic *akrasia* we described earlier. The latter requires us to believe where we judge that belief is inappropriate. Some cases of wishful thinking may take this form, but it may also arise when we believe what we *ought* to judge to be ill supported evidentially and thus normatively inappropriate. Indeed some cases of wishful thinking appear to depend upon our being ignorant of (or deceived about) the warrant our belief possesses. Full-blooded epistemic *akrasia*, if it is to be found, should lack this dependence upon (culpable?) ignorance or (self?) deception. It seems important that at least some *akrasia* be distinct from self-deception. Moreover, if practical *akrasia* normally involves a background of self-deceived belief, then putative epistemic *akrasia* may simply collapse into self-deception. We now turn to a consideration of how far this is the case. Can full-blooded epistemic *akrasia* be understood as a special form of the *akrasia* of inquiry?

Before doing this, however, it will be useful to make two observations about the structure of inquiry and deliberation and the norms that govern these activities. One point emerges from our discussion of ordinary practical *akrasia* and has already been alluded to here: There will be an important interplay of evaluations, evaluations all things considered, general resolutions and commitments, and more specific commitments. An adequate account of epistemic activities must take note of the special kinds of evaluations and commitments that they involve. My practice may be affected by evaluations and commitments that are not, properly speaking, epistemic. I may attach great value to sharing the religious beliefs of those with whom I must live. Or I may acquire a moral commitment never to think of anyone as wholly evil. Each of these may require me to shut my eyes to the weight of evidence or to the reliability of the methods that I employ. "All things considered," I may judge, I should hold a belief which is poorly supported by evidence or which was formed in an unreliable way. If I remain agnostic in spite of my values, or if I conclude that some individual is, indeed, truly evil, this may be a failing, but it is not an epistemic one.

However, if I succeed in retaining my faith, or if I succeed in identifying a germ of humanity in Adolf Hitler, this may fail to conform to my epistemic standards but, "all things considered," it need not be a failing. In a broad sense, the resulting belief need not be "normatively inappropriate." Indeed, we may even imagine cases of (non-epistemic) *akrasia* which occur because someone cannot help being too assiduous in apportioning belief to evidence: In spite of her believing that evidential considerations are not decisive in connection with religious belief, she finds herself unable to avoid a kind of scientistic atheism; or in spite of believing that one should be loyal and trusting of one's friends, she finds herself making a dispassionate assessment of the evidence when one of them is accused of a crime.

This makes it difficult to specify exactly what would be required for a case of full-blooded epistemic *akrasia*. I suggest that we adopt the following, probably oversimplified, picture. We shall restrict attention to inquiries and deliberations that are governed by an overarching commitment to solving a problem or assessing a belief relying solely upon considerations that are relevant to truth. These are activities that have goals that are fully cognitive. Then we can define epistemic *akrasia* as a distinctive form of irrationality which is internal to these "fully cognitive" inquiries. We employ means whose use is, we are fully aware, inconsistent with the values and commitments which apply to fully cognitive inquiries of this kind, or which emerge, rationally, within this particular inquiry.[3]

Now for the second point about the structure of inquiry, which concerns the role in it of questions and questioning. An inquiry is an attempt to solve a problem or, most commonly, to answer a question: it succeeds when we arrive at a solution or answer which meets the commitments that govern the inquiry; in the case of fully cognitive inquiries, when we arrive at an answer that is true. If inquiry is at all reflective, then we shall face further problems, pose further questions, as it proceeds. We can question the methods employed, the methods used to evaluate those methods. We can ask whether further information is available (or even already possessed) that is relevant to the success of the inquiry. We can put questions to others in order to get their opinions of methods or in order to elicit information that they possess, and so on. The success of inquiry depends on whether we ask the *right* questions. Things can go wrong if we fail to raise important and relevant issues, and they can also go wrong if we ask too many questions. The overcautious are likely to ask too many questions; the credulous generally ask too few. Our mastery of fundamental epistemic norms is manifested in the questions we raise and, just as important, in the questions we don't raise. The norms thus often have a negative character: They are reflected as much in facts about what does not occur to us as in the rules we formulate and reflectively follow.[4]

4. Why Does Epistemic *Akrasia* Seem Problematic?

There is one big difference between full-blooded *akrasia* and some forms of practical *akrasia*. Even if forming a belief is an action, beliefs, unlike actions, are not data-

ble events. Rather, they are enduring states of people. Beliefs are more like resolutions and intentions than they are like actions: When I accept a proposition, I acquire a commitment to plan my actions and deliberations on the assumption that this proposition is true. Epistemic *akrasia* will thus involve conflict between commitments and evaluations or between more general and more specific commitments. For ease of reference, we shall distinguish the conflicting elements as (a) the belief and (b) the normative component. The latter may be a cognitive commitment or an evaluation. The aim of this section is to investigate why full-blooded epistemic *akrasia* can seem problematic and to identify the questionable assumptions on which this appearance rests.[5]

To avoid misunderstanding, I should distinguish the case that interests me here from some others. I am not concerned with cases where the initial belief is held unconsciously and thus does not feature in reflective deliberation. Nor am I concerned with cases where, although both components are accessible to consciousness, only one is actually "accessed." Nor am I concerned with cases in which it is experienced as an irrationally strong tendency to believe, which the agent succeeds in resisting and from which she feels constantly alienated. Nor am I concerned with a case where the mother believes that it is right, in the circumstances, to maintain her son's innocence in the face of the evidence. A genuine case of full-blooded *akrasia* would have the following components: Both the belief and the normative commitment are "present" to the agent, and in some manner, she is aware of each.

- She has a genuine commitment to each component.
- She is aware of the conflict between their different demands.
- She is aware that she is committed to eliminating this conflict in a different way from that which she actually employs.

It is best to work with an example. Consider a mother who believes that her son is innocent of some particularly heinous crime of which he has been accused. For epistemic *akrasia* to be possible, she must intend her belief to be fixed by the balance of the evidence — her inquiry is fully cognitive — and her state must have a normative component that renders her belief inconsistent with this intention. This may consist in accepting one of the following:[6]

- The evidence supporting her son's innocence is slight.
- The evidence available to her is too limited to support a judgment either way.
- There are strong or conclusive reasons to believe in her son's guilt.

Each component of her cognitive position is present to consciousness. She can be aware of her commitment to both without immediately losing her attachment to either, in spite of the fact that she is aware that they are epistemically inconsistent. A common worry is that this is not possible, that (for example) her confident belief in her son's innocence will be sufficient to cast doubt on the sincerity of her endorsement of the other propositions, or her normative judgment will inevitably under-

mine the original belief.[7] The remainder of this section will explore some considerations that support this natural view.

So long as the activation of these conflicting beliefs is temporally distinct, there need not be a problem. Whenever she is in her son's presence, one set of sentiments ensures that she trusts his avowals of innocence and is skeptical of the evidence that supported her earlier normative belief. But when closeted in her lawyer's office, another set of concerns ensures that these avowals seem shallow and insincere once she confronts the weight of evidence that counts against him. Each belief fills its explanatory role while the influence of the other is temporally suppressed. While this may be irrationality, it is not full-blooded epistemic *akrasia*. The latter requires both beliefs to be accessible — and indeed accessed — at much the same time as part of a single process of inquiry or deliberation. The agent must be aware of the force of the one, even as she acts on the other.[8]

We shall start with something utterly uncontroversial. We often appeal to people's beliefs in order to explain their outward behavior and other features of their mental lives. If I know that someone believes that the library closes at six o'clock, I can understand why he rushes toward the library when he notices that it is already five fifty. I shall also understand why he spends time wondering whether he can get from his office to the library in less than ten minutes, and also the irritation he feels when he decides that he cannot, or his surprise on seeing that it is still open at seven. I can also explain why he says "six o'clock" when asked when he thinks the library will close. Supplemented with other information about an agent's desires and attitudes, beliefs can be used to explain how the agent behaves, the course of his reasoning or deliberation, his emotional and other affective states, and his avowals. Such explanations work only if the agent would have acted, reasoned, or felt differently had his belief been different.

There is something unsatisfactory about the idea of a "genuine" belief or commitment that can *only* be manifested in an agent's avowals, that is insulated from all the other kinds of manifestation that I mentioned. If this was possible, then full-blooded *akrasia* would be wholly unproblematic. Someone could avow that it was epistemically wrong to believe some proposition, while belief in that very proposition would feature in the best explanations of her external behavior, her deliberations, and her feelings. I shall simply assume that if someone avows belief in a proposition when she has absolutely no inclination to use that proposition in planning her conduct or in evaluating her beliefs, and who also has no inclination to feel embarrassment, shame, or surprise at this fact, then her avowal, even if sincere, is false. Thus, a further necessary condition for full-blooded epistemic *akrasia* is that neither of the conflicting states should be insulated from exercising all of its causal propensities. If they are genuine mental states of the agent, they can contribute to explanations of her behavior and of other features of her mental life. We can formulate this as a relatively weak principle:

P1. That *X* believes that *p* cannot be made true solely by the fact that *X* candidly asserts or endorses either the proposition that *p* or the proposition that she believes that *p*.

How should this lead to doubts about the possibility of full-blooded *akrasia*? Problems might arise if the causal explanatory propensities associated with the two conflicting states ensured that one state could exercise its causal propensities only if the other did not. As is suggested by their role in explanation, suppose that ascriptions of commitments to people support subjunctive conditionals. It is plainly impossible that the following two conditionals be true in a situation where the antecedent of each was satisfied, where the agent both believed that *p* and held normative commitment *N*.

1. If *X* believed that *p* and was in situation *S*, then *X* would do *A*.
2. If *X* had normative commitment *N* and was in situation *S*, then *X* would not do *A*.

Normally our first-order beliefs and our beliefs about what it is rational to believe are in harmony: We believe what we think we ought to believe. And it seems plausible that it would make little sense to think of someone as a subject of beliefs if their first-order beliefs and their beliefs about what it is rational to believe were never in harmony, or, indeed, if it were not generally the case that they were in harmony. In cases of full-blooded *akrasia*, these connections are broken. If our subject both believes in her son's innocence and believes that the weight of reasons supports his guilt, we need to understand how the mother's belief in her son's innocence, and her commitment to the irrationality of such a belief, can simultaneously possess an appropriate role in cognition without giving rise to any conflicting conditionals of the kind we have just described.

Let us make one further assumption:

P2: If *X* is a reflective rational agent, then, in general, she can control her actions and opinions by reflecting on what there is reason to believe or do.

This requires that our reflective beliefs, including beliefs about what we believe and beliefs about what we *ought* to believe, stand in reasonably stable causal relations with our other beliefs and our practical and theoretical decisions. By and large, we can suppose, our beliefs and actions reflect our (assessment of our) reasons. When we decide that one of our beliefs is irrationally held, then, ceteris paribus, we lose confidence in it. When we decide that, all things considered, we would be rational to perform some action, we are likely to do so. One aspect of this may be that our deliberations are sensitive to the demands of reason, and our actions and beliefs are sensitive to the routes taken by our reflections and deliberations.

Let us return to the example of the mother and the son and, provisionally, make a surprisingly common assumption: Our beliefs (in conjunction with desires and other attitudes) are primarily manifested in our *action,* rather than in patterns of deliberation and feeling. Suppose now that the mother has the goal of preserving the reputation of her family. She has sufficient reason to adopt this goal and has reasonable views about the relative priority of her different goals. Suppose she also believes the following:

1. If her son is innocent, the reputation of the family is best preserved by declaring her belief in his innocence and doing all she can to secure his acquittal.
2. If her son is guilty, the reputation of the family is best preserved by denouncing his immorality, disinheriting him, and announcing that she no longer sees him as a son of hers.

The occasion has now arisen when she must decide whether to stand by her son or denounce him. When she works out how to act on this occasion, what would we expect her to do?

We can get a sense of paradox as follows. If we take account of her belief in his innocence and also recall that her goal is to preserve the reputation of the family, then would expect her to stand by her son. But, given that she believes that it is *rational* to believe in his guilt, she may also believe that it would be rational for her to denounce him. In line with P2, she may well approach her decision reflectively, wondering which action there is better reason to perform. And she may accept that if there is reason to believe in his guilt and also reason to believe that if he is guilty, he should be denounced, then she has good reason to denounce him. Thus, even if she believes in his innocence, it can still follow that if she is reflective in planning her actions; then she will act as she should if she believes in his guilt. Reflection can always cut short the expected effects of her akratic belief upon her behavior. It begins to look as if we have conflicting subjunctive conditionals of the problematic kind:

1. If she believes that her son is innocent, she will defend his reputation.
2. If she believes it is rational to believe that her son is guilty, she will denounce him.

If she is generally reflective, and thus generally acts as if she believed in her son's guilt, it is hard to see what her belief in his innocence can consist in. The pattern in behavior naturally associated with belief in guilt is present, albeit produced by the role in inference of the apparently distinct belief that there is good reason to believe in his guilt. If, on the other hand, she behaves in accordance with her belief in the son's innocence, the belief comprising the normative component appears to be explanatorily inert, in which case it is hard to see what makes it that she has *this* belief. We may then begin to wonder whether the first-order belief and the belief about what it would be rational to believe are genuinely "distinct existences." And in that case, epistemic *akrasia* seems highly problematic.[9]

A simple example of this kind is insufficient to establish that epistemic *akrasia* is genuinely problematic. However it does give rise to a significant challenge. If full-blooded epistemic *akrasia* is possible, then we must give an account of how the belief and then normative commitment can each be operative in the agent's cognitive life at the same time. We must be able to trace their manifestations back to the beliefs that explain them. If the mother denounces her son, what makes it the case that she does this because she thinks it rational to believe in his guilt rather than because she believes in his guilt. If she continues to defend him, what is there apart from her avowal to show that she retains her commitment to the wrongness of her belief.

What do we learn from this example? Full-blooded epistemic *akrasia* requires that both the belief and the conflicting normative commitment should be operative at the same time. (The akratic smoker is aware of acting on her desire for a cigarette while fully conscious of her commitment to stopping smoking.) We need an account of how these beliefs and commitments can be "manifested" or "operative" that allows that what counts as a manifestation of the one does not count against the reality of the other. The example suggested that if we restrict our attention to manifestations in external behavior, it might be difficult to give such an account. The public manifestations of believing that it is rational to believe that *p* may be no different from the public manifestations of believing that it is rational to believe that *p*. Hence we shall only understand how epistemic *akrasia* is possible by looking at a wider range of manifestations of beliefs and commitments. At the beginning of this section, I noted that as well as contributing to the explanation of our behavior and avowals, beliefs and commitments could be invoked to explain both the routes taken by our reasoning and deliberation and our feelings and emotions. And when we described the two ways in the which the mother could arrive at different answers to the question how she should treat her son, we paid attention to the different ways in which her deliberations could go. We shall see in the next section that attention to the process of deliberation and to feelings are both required if we are to make sense of the distinctive roles of our first-order beliefs and our normative commitments. This will enable us to see how epistemic *akrasia* is possible.

5. How Epistemic *Akrasia* Is Possible

One lesson of the example we have been using is that we should not fix the functional role of (conscious) beliefs by reference to broad patterns in belief, desire, and behavior. How the mother will act depends on how she reflects, on the routes taken by her deliberations and inquiries. Our conscious beliefs provide premises for reasoning, and similar patterns of beliefs can produce different conclusions and different actions according to how they are deployed in processes of reasoning and reflection. Not only must we attend to the ways in which beliefs guide deliberation, but (a second lesson) we must take account of how our beliefs are activated, of when we take note of them and admit them to processes of deliberation. If she is extremely reflective, then, as we have seen, her beliefs about what it is rational to believe may determine her action and her first-order belief may not be "activated"; if her reflections take different directions, then the first-order belief may influence her choice while the second-order belief does not. Hence we must take account of how and when beliefs are activated in deliberation.

Let us work with a somewhat simplified picture of how beliefs enter our deliberations. One way in which our beliefs are activated, summoned to play part in our deliberations, is through our posing questions to which these beliefs provide our answers. This parallels the way in which our beliefs can be called on in cooperative inquiry: Someone asks the question and we give our answer in a form appropriate to the current state of the conversation and inquiry. In the cooperative case, a piece of

information possessed by one of the cooperating agents may fail to influence the up-shot of the inquiry if the question required to elicit that piece of information is never asked, if it never becomes salient. Cooperative inquiry can fail because one partici-pant fails to ask the right questions of the others. And it can fail because another par-ticipant fails to point out that some relevant question has not been raised. Perhaps a parallel phenomenon is found in solitary deliberation. If the question whether p does not become salient, then my belief that p may not engage with my deliberations. In an earlier essay I expressed this point by saying that self-questioning provides a process by which propositions can be elicited from our store of information (Hook-way 1997).[10]

So our subject's two beliefs may be elicited through her (or someone else) raising the questions:

Is it the case that p?
Is there better reason to believe that p than *not-p*?

It seems evident that these are *different* questions. We can see that in many cases they have different correct answers. Our concern is with the possible relations between the answers I *take* them to have, and with the effects of which of these questions I ask on how I act.

The following two things seems clear:

1. I can raise and address the second question even if I have no settled answer to the first. Indeed, the second can become salient simply because I am currently agnostic about the first matter.
2. Cases in which I raise the first question can be divided into two sorts: (a) I may just candidly offer the answer I happen to have stored away. (b) I may be prompted to ask the second question as a way of checking my answer to the first one, or in order to replace my current agnosticism about the first ques-tion with a firm opinion.

The current proposal is that we can compare different routes that reflection can take by comparing the different questions that are raised (reflectively or unreflectively) as it proceeds. These questions can be used to set the targets of our deliberation, to identify the practical and theoretical problems we aim to solve. They can be raised as reflective comments on, or challenges to, the progress of these deliberations, and they can also be used to elicit or activate beliefs or items of information that are al-ready possessed by the inquirer. The information I possess, my current beliefs, will have an impact on the progress of the deliberation, according to this over-simplified picture, only if a question is raised to which the belief provides my answer.

To summarize this part of the discussion, reflection, including both practical and theoretical deliberation, is an activity that can be controlled through the exercise of normative standards. Where it ends up will depend on what we *attend* to, on what we *notice*, on which of our beliefs is *activated* in the course of our reflections. That a belief may fail to be salient—fail to be activated—when it is relevant to matters

under discussion may be a failure of rationality: The belief may be accessible, even if it is not accessed. One element in becoming rational is learning to— being trained to— ask the right questions. We must ask the right critical questions if we are to expose errors in our reasoning; and we must also ask the questions that lead us to access our beliefs when they are relevant to our deliberative concerns.

This suggests one necessary condition for the intelligibility of full-blooded epistemic *akrasia*:

> It must be possible to raise the question *p* without at the same time raising the question whether belief in *p* meets some prescribed normative standard. These are different questions.

Another necessary condition is:

> It is possible for someone's candid answers to these questions to be different.

It would be possible for this to be the case yet, either due to self-deception or to the fact that the times in which the different questions are raised are significantly different, the agent does not (perhaps even cannot) notice that this is the case. A further necessary condition for the possibility of epistemic *akrasia* is that this conflict is available to, or perhaps "is noticed by," the agent. The agent can be aware that her candid answers to these two questions are in tension; and this can incur within a process of deliberation for which, according to the agent, both questions are relevant.

Although these are necessary conditions for epistemic *akrasia*, they are by no means sufficient. Recognizing such tensions can promote their removal: The mother may be led to reassess her candid assurance of her son's innocence; or she may reexamine the grounds of the normative judgment, confident that they will be found to contain errors; or her confidence in both judgments may be dramatically reduced. *Akrasia* requires that she acts on the basis of her judgment of her son's innocence while, at the same time, continuing to endorse the normative claim that all the evidence confirms his guilt. The challenges that this presents are twofold: We must arrive at a satisfying description of the phenomena, one that makes it plausible that they should occur; and, in the light of the previous section, we should explain how both commitments continue to be operative in what is going on and also why the agent has been guided by the one which, she holds, should *not* exercise authority over the other. The remainder of this section will address the first of these tasks.

In our discussion of the example of the mother and son earlier in this section, we considered two deliberative routes, one leading through beliefs in her son's innocence to her public defense of her honor, and the other leading through beliefs about the rationality of belief in his guilt to an act of public denunciation. In the case we are considering, her deliberation could involve rehearsing each of these argumentative routes, perhaps successively examining first one and then the other, oscillating back and forth between the two possibilities, each powerless to silence the other. How she acts may depend on the stage of this oscillation where she decides to terminate her

deliberations and act on the belief which is, at that moment, most salient. It is influenced by which questions she chooses to ask last: the irrationality displayed is thus a form of practical irrationality. Whichever wins out, she can respond emotionally. Given her initial commitment to a fully cognitive inquiry into the matter, she will be saddened if her judgments of rationality win through and she has to accept the likelihood of her son's guilt. But if her belief in the son's innocence wins, she will experience a variety of reactive emotions, perhaps shame or guilt. In some cases, the second argument will continue to reassert itself and she will understand how her inability to sustain her commitment to a fully cognitive inquiry has led her to act in ways that she will regret. If it does not, she may continue to be aware that her confident belief may depend on irrelevant determinants of when her deliberations should be brought to a close. Perhaps the success of the belief will cause the normative commitment to fade, but, once again, there is no reason why she cannot be aware that this is so.

This is, of course, an oversimplified description of a particularly vivid instance of epistemic *akrasia*. It embodies the following elements. First, we can allow that each commitment may be manifested unproblematically in other deliberations in which its rival has no role. Second, victory in the deliberation we are concerned with here results from processes that are in tension with the mother's commitment to a fully cognitive inquiry. Third, the losing commitment need not simply disappear but can continue to be manifested in the reactive emotions that accompany the mother's reflections on her beliefs and her actions. And fourth, it acknowledges that we can make sense of these phenomena only by attending to the conduct of activities such as deliberation and inquiry.[11]

6. Conflicts of Intuition: Some More Examples of *Akrasia*

An important step in the argument of the last section was the recognition that the reactive emotions of the mother can attest to the presence in her deliberations of commitments that were somehow silenced or defeated when she decided how to act. This is not the only way in which affective states have a role in the regulation of deliberations, in which they can register the presence of standards of evaluation that are not consciously articulated or acknowledged. In an earlier essay, I argued that we can make sense of the epistemic role of states of *doubt* when we notice that they generally involve a motivational component in the form of an anxiety about the agent's grasp of the proposition doubted (Hookway 1998). Something similar is likely to be involved in the case that we described. One piece of evidence that the defeated normative standards are operative at the moment of the mother's decision is that she will feel anxiety as she acts on her belief in her son's innocence. This affective acknowledgment of the completing claim betrays sensitivity to the irrationality of what she is doing, a sense that her actions may go awry through her own fault.[12] Describing it as "anxiety" explains its "motivational" role: it can stimulate us to reflection, or to abandoning inferences and judgments that are contrary to such intuitive

responses. Thus epistemic anxiety can be an immediate unreflective acknowledgment that the deliberator is in danger of going wrong.

Such anxieties can, of course, be weak and they are not always rational. But it is plausible that what philosophers often describe as "intuitions," — the intuition that an argument is not a good one, that a concept does not apply to a particular situation, that a sentence is syntactically out of order — are immediate affective embodiments of norms that we follow but which we cannot explicitly formulate. We express anxiety about accepting these claims without fully knowing why or how. Rationality involves trusting, or listening to, our "intuitions." A distinctive form of *akrasia* — quite a full-blooded one — can come from the motivated refusal to listen to our intuitions. The intuition provides our only conscious access to a normative commitment by signaling when we are contravening it. Thus we can be aware that our beliefs or inquiries are in conflict with commitments that we cannot formulate or acknowledge. This can illuminate some familiar phenomena of irrationality.

Familiar psychological studies of reasoning suggest that humans are naturally inclined to accept a variety of inferences that are evidently irrational: familiar examples concern probabilities or inferences that turn on the understanding of conditionals. Getting it right, in such cases, can be hard work, requiring us to resist kinds of inference that are ingrained by habit or built into our cognitive architecture. Those of us who have read the textbooks know that we are likely to go wrong in these cases and struggle to do better. This does not prevent inferences that we know intellectually to be flawed from feeling compelling. The temptations to think that a run of reds raises the probability of the next spin being red, or to conclude that Linda is more likely to be a feminist bank teller than to be a bank teller, do not simply fade away. Experience may warn us of the likelihood of error in such cases. Although we are naturally drawn into the inference, we encounter the inchoate warning that something is wrong and we should attend more carefully to what we are doing. There are two intuitions here: that the inference is a good one, and that such intuitions are not always to be trusted. If we are committed to fully cognitive deliberation, we know which one should be trusted. Rationality requires us to manage these intuitions, to understand their sources and form our beliefs as we should. In each case, we can be more strongly motivated to go the wrong way, to go with an argument that seems intuitively right when we intuitively know that this intuition is untrustworthy. The normative component of this form of *akrasia* is thus an affective commitment to normative requirements rather than a full belief. But this does not prevent regret, shame, and anxiety from being present just as in the case described in the last section.

7. Epistemic Virtues

In section one, I suggested that our examination of these phenomena of epistemic *akrasia* lends support to the view that effective cognition depended on possessions of states such as virtues, enduring states of character with a role in regulating inquiries and ensuring their success. The examples we have considered have illustrated some

ways in which cognition can go wrong; and effective responsible inquiry depends on the mastery of epistemic norms that can prevent it going wrong in these ways. The discussion has relied on a particular picture of epistemic evaluation. Its primary role lies in the regulation of activities of inquiry and deliberation. These activities require work: We often have to make an effort to carry them out effectively. We actively address problems and try to formulate and answer questions. We compare our different beliefs and hypotheses, trying to bring to bear opinions about the weight of evidence on propositions we accept or those about which we are agnostic. Success will depend on our doing the right amount of work: addressing issues that need to be addressed without wasting time and energy over irrelevant matters. It also depends on our efforts having their expected effects. If we spend time collecting and assessing the evidence that supports some proposition, then our results should be reflected in the degree of support we give the proposition. It is easy to see how indolence, partiality, prejudice, and a range of other interfering factors can prevent our making our efforts appropriate or prevent them from being effective. Our epistemic values should equip us to carry out a managerial task: employing our epistemic resources and exercising our epistemic efforts in ways that enable us to solve problems and reach true answers to significant questions.

The present essay has identified some of the tasks that effective regulation of deliberation faces, along with some of the potential conflicts we must be able to deal with. We shall now list and elaborate some of these.

- When normative assessments of evidence and argument conflict with confidently held beliefs, a judgment must be made concerning which should have authority over the other.
- Our commitment to fully cognitive inquiry must contend with (often unconscious) inclinations to favor one solution over another by focusing on questions that send our inquiries down potentially distorting routes. This can lead us to abandon the commitment or it can prevent our seeing that it is not fully effective.
- We must be able to weigh the force of apparently incommensurable evaluations, for example, formulated commitments against contrary "intuitive" anxieties.
- We must be able to weigh the force of conflicting evaluative intuitions. In these cases we may not be reflectively aware of the normative standards that are reflected in the intuitions. It may not even be evident whether they really conflict.
- We must be able to deal with the fact that our reflective normative judgments may simply be powerless to adjust our beliefs as rationality requires. This may be due to acquired habits, to properties of our cognitive architecture, to brain damage, to laziness and inattention, to emotional attachments, and to a range of other causes.

These kinds of phenomena draw attention to two respects in which our deliberations are not under our control. First, much depends on whether we raise the right reflective questions. Our mastery of norms is reflected in the questions we don't raise

as well as in the questions we do raise: we are sensitive to irrelevance as well as to relevance. As we have seen, such norms have a negative character: it would be hopeless if we have to consider every possible question and explicitly apply rules in order to decide that it was irrelevant to our deliberations. Second, we cannot control whether the answers we reach to our reflective critical questions will have their intended to effect. The recognition that some belief is poorly grounded may simply be impotent to shake confidence in the belief unless appropriate mechanisms are in place. And the factors that shape our choice of questions, like the factors that influence the effects of normative judgments on their doxastic objects, are not open to introspection. Much of the time, we don't know what is going on or why.

If this is right, then effective epistemic agency depends on the possession of *confidence* in our intuitive judgments about which questions should be asked and about the relative weights of apparently incommensurable evaluations. Unless we possess this confidence, we will constantly face questions of how to proceed that we are powerless to answer. This is because intuitions (for example momentary states of anxiety) often provide our best access to normative standards that are ours, but which we are unable to articulate. Unless we can trust these intuitions—and our intuitions about when they can be trusted—those normative standards cannot influence our cognitive activities. Effective epistemic agency also requires that this confidence is not misplaced. Trusting our intuitive judgments ensures that we *do* ask the right questions and that our comparisons of different evaluations do not impede our cognitive projects. The confidence provides the *internalist* dimension, and its not being misplaced constitutes an *externalist* dimension, of epistemic evaluation.

If we possess this confidence, and it is not misplaced, then our deliberations and inquiries will be broadly "continent." It is natural to think of this as the possession of a virtue: a state of character that ensures that we take heed of acknowledged reasons and maintain our rational commitments. It may be best to think of this as based on a cluster of capacities and skills. Some may be innate; others the result of training and education; yet others the product of conscious thought and planning. They are unified by their common role in the evaluative practice that regulates inquiry and deliberation. I am not committed to claiming that there is a single mechanism that does the whole job; nor that if any members of the cluster are present, then all must be; nor that they need be unified by a common location in the accounts of the mind produced by cognitive psychologists.[13]

If "continence"' is a virtue then, like *courage*, it is what has been called an executive virtue (Pears 1978, Wiggins 1987) or a "virtue of will power" (Roberts 1984: 230). Such virtues are exercised for the sake of some further end: success in inquiry or victory in battle. As Roberts puts it, they are concerned with "self-management" or self-mastery, with the ways in which we cope with the different motivational pressures we face and plan our actions in the light of these. Thus courage would be an example: It is a virtue that can be deployed in the service of any of a variety of ethical (and epistemic) outlooks. While I may be benevolent or temperate for its own sake, I cannot act merely for the sake of courage. The function of such virtues is to enable us to overcome the obstacles to achieving the further goal—to escape from the inclination to avoid the risk of being maimed or killed in battle, or to escape the

risks of lapsing into wishful thinking or procrastination and thus thwarting our deliberative aims in the case of continence. Irrationality, in general, is a threat to our cognitive aims, and the cluster of dispositions and capacities provide the knowledge and motivation to do the work that is required to avoid it having this effect. It is plausible to describe continence as a vehicle of self-control. It helps to ensure that *my* commitments and values shape the development of my opinions. And it ensures that *my* commitments endure unless it becomes rational to reassess them.

A true Aristotelian would be uneasy about linking virtue to self-control. The latter appears to involve an ability to recognize contrary impulses and inclinations and the power to resist them, to face them down. For Aristotle, this is a response to a difficulty that the truly virtuous person would not face. If we are truly virtuous, then contrary impulses and inclinations do not interfere with the comfortably virtuous life: Continence is a way to cope with the fact that we are not ideally virtuous.

However it is still appropriate to describe these states as virtues. Continence is an enduring state of a person that enables him or her to deliberate well. It does this by exploiting a body of normative standards and capacities for evaluation. The standards of evaluation employed are generally not (and probably cannot be) formulated as precise formal rules. In many cases, application of rules to particular cases will involve weighing apparently incommensurable values and arriving at an intuitive judgment the bases of which are not fully explicit. Finally a state such as continence has a role in motivation: it enables the agent to inquire well, to adjust beliefs automatically in the light of normative considerations when she judges that it is appropriate to do so. These, I take it, are all marks of a state being a virtue.

A brief illustration of the point about judgment may help here. In listing the capacities we require for effective inquiry and deliberation, we mentioned the ability to weigh apparently incommensurable epistemic values. Suppose it is one of our epistemic duties to subject testimony to suitable scrutiny before accepting it; we should avoid gullibility so far as is possible. This requirement has a prima facie character: We should respect it so long as more pressing epistemic requirements do not conflict with it. Where obtaining false testimony carries few risks, or where it is important that our cooperative investigation advances quickly, or where the source of testimony is a trusted colleague and her testimony deals with a topic where there has been no reason to doubt her reliability: In all these cases, it would be best if the question of reliability was not made the subject of any reflection or investigation. Deciding whether this is the case where the "duty" calls for epistemic action involves comparing the weight of different epistemic desiderata where we lack any overarching rule that tells us how to decide. When we bear in mind the complexity of any particular case, and the need to apply our duties to particular cases, it is easy to see that we rely upon experience and *judgment.* Even if we decide that this is a case where the informant's reliability should be assessed, there will be many ways of conforming to this duty. We might recall her past record, collect the opinions of others, test her in areas where we already possess reliable information, and so on. How scrupulous we should be is a matter of judgment, again a way of weighing a variety of epistemic and non-epistemic goods. Suppose we then obtain different sorts of evidence that have a bearing on her reliability, some positive and some negative. Once again we

must weigh them. And once again we are unlikely to have formal rules that can guide us in doing so. Rules and duties must be applied to complex cases in the course of well-regulated inquiries and deliberations. And this cannot be governed by further explicit rules, on pain of a regress (see Larmore 1987, chapter 1).

It is easy then to see how our examples of epistemic *akrasia* involve failures of virtue. In some cases, the agent's judgment is deficient: She attaches too little weight to normative considerations that she is committed to taking more seriously than she does. In others, the failures are ones of motivation: her anxiety as she acts on her belief in her son's innocence is a sign that a normative commitment that she endorsed, and that she is still committed to giving authority over her beliefs, has failed to have its due effect upon her current cognitive state.[14]

Notes

1. This argument could provide part of a defense of virtue epistemology. It would need to be supplemented, inter alia, by a demonstration that the standards that regulate inquiries and deliberations are the most fundamental epistemic norms. This contrasts with the more common epistemological view that the evaluation of states such as beliefs as justified or as knowledge is the core of epistemology. The second stage would consist in showing either (a) that the latter concepts can be explained by reference to notions of good inquiry or theoretical deliberation, or (b) that the concepts of knowledge and justified belief are less central, less important for understanding the pursuit of truth, than is often supposed. The current essay is not concerned with this further step. Even without this second stage, the argument should establish that epistemic virtues are more than just useful bodies of habits that enable us to achieve more automatically and easily what would otherwise place great demands on our attention and our reflective powers. Without such virtues, I shall argue, rational deliberation and inquiry would be impossible.

2. Examples of this kind illustrate what Amelie Rorty (1983) has called "*akrasia* of inquiry." Someone may succumb to ways of thinking she knows to be intellectually flawed—for example, being unfair to the ideas of others or accepting something on the authority of someone she knows not to be trusted (Zagzebski 1996: 154–55). Section II of Rorty's essay provides a useful taxonomy of kinds of "doxastic *akrasia*": most of them seem to fit the pattern just described. What I have called full-blooded epistemic *akrasia* is discussed more fully by Heil (1984).

3. This strategy can be generalized to a wider range of cases. Even if an inquiry has goals that are not fully cognitive, it may comprise subinquiries that are taken to be fully cognitive. Even someone who wishes to form religious beliefs on grounds that are not fully cognitive can display epistemic *akrasia* when he investigates just what the prevailing local beliefs are.

4. This topic is explored more fully in Hookway (1999).

5. One of the earliest attempts to undermine the common assumption that epistemic *akrasia* is impossible is in Graham (1974).

6. John Heil expresses this by saying that the akratic believes two propositions that are 'epistemically incompatible': accepting one of them would constitute a reason for not accepting the other (1984: 63).

7. This suspicion is clearly expressed by Lear's comment that the incontinent is 'a stranger to himself' and that 'it is in his actions, not in his assertions, that he may discover who he is' (1988: 186).

8. This diagnosis is similar to that of T. M. Scanlon who holds that the source of doubt

about the possibility of epistemic *akrasia* 'is the idea that judging *P* to be supported by the best evidence is so immediately connected with believing *P* that there is no room for slippage of the kind that can occur between judgment and action' (1999: 35). Scanlon's rebuttal of this claim identifies cases where someone is guided in believing a proposition by a consideration which, at the time, *seems* to be a good reason for belief although she is aware that other considerations establish that it is not a good reason at all (36).

9. These considerations seem to support Linda Zagzebski's contention that "intellectual *akrasia* involves self-deception more than does moral *akrasia* because there is probably a stronger link between believing and believing justified than doing and believing right" (1996: 154).

10. We should guard against a misunderstanding here. I am not claiming that *in general* beliefs influence behavior only when they are explicitly activated by self-directed questions. Dispositional beliefs can ground habits of inference or influence the ways in which we describe our experience of finding things salient. The claim is that some beliefs, on some occasions, play a role in the formation of behavior through being elicited to serve as premises in conscious inference. It is only beliefs that can be invoked in this way that can be avowed and made a matter of conscious reflection. Full-blooded *akrasia* can occur when beliefs of this kind coincide with coexisting beliefs of the same kind that express a negative evaluation of their status. Phenomena that are close cousins of such *akrasia* may occur when just one of the beliefs in question is available for conscious elicitation.

11. Some philosophers interested in these phenomena think we should divide the self or introduce homunculi. That does not seem to be required under the current proposal. Rather, it is unproblematic that the questions to which these different propositions are answers need not be asked together if they are asked at all. In that case it is unproblematic that the different "beliefs" can be activated at different times and in different ways by asking appropriate questions.

12. As Mark Johnston (1988) has noted, some forms of irrationality (for example, wishful thinking and self-deception) may stem from a reluctance to face up to anxieties about ways in which more careful investigation may thwart our aims or force us to confront unpleasant truths. He sees self-deception as often involving a kind of intellectual cowardice. It is possible that we should view continence as a special application of courage to the case of inquiry and deliberation—or see both as different forms of a single, more abstract virtue. But I cannot pursue that issue here.

13. Pace Harman 1999.

14. In producing the final version of this essay, I have greatly benefited from the discussion at the conference in Santa Barbara. Wayne Riggs was commentator on that occasion, and his reactions, both at the conference and subsequently, have led to many improvements. I am also grateful to Lucy Burroughs, Jonathan Dancy, Tobies Grimaltos, Stephen Makin, David Owens, Jenny Saul, and Leif Wenar for very helpful discussions of drafts of the essay; and to audiences at talks on this material delivered at the Universities of Bristol and Manchester.

References

Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson Donald. 1985. "Deception and Division," in *Actions and Events*. E. Lepore and B. McLaughlin, eds. Oxford: Blackwell, 138–48.

De Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge: MIT Press.

Graham, Keith. 1974. "The Limits of Irrationality." *Inquiry*, 17: 315–26.

Harman, Gilbert. 1999. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society*, 99: 315–31.

Heil, John. 1984. "Doxastic Incontinence," *Mind*, 93: 56–70.

Holton, Richard. 1999. "Intention and Weakness of Will," *Journal of Philosophy*, 94: 241–62.

Hookway, Christopher. 1997. "Analyticity, linguistic rules and epistemic evaluation," in *Thought and Language*. J. Preston, ed. Cambridge: Cambridge University Press, 197–218.

———. 1998. *Truth, Rationality and Pragmatism: Themes from Peirce*. Oxford: Oxford University Press.

———. 1998. "Doubt: affective states and the regulation of inquiry," in *Pragmatism*, *Canadian Journal of Philosophy, Supplementary Volume*. Cheryl Misak, ed. 24: 203–26. Reprinted as chapter 10 of Hookway (2000).

———. 1999. "Epistemic Norms and Theoretical Deliberation," in *Ratio*, 12: 380–97.

Johnston, Mark. 1988. "Self-Deception and the Nature of Mind," in *Perspectives on Self-Deception*. B. McLaughlin and A. Rorty, eds. Berkeley and Los Angeles: University of California Press, 63–92.

Larmore, Charles. 1987. *Patterns of Moral Complexity*. Cambridge: Cambridge University Press.

Lear, Jonathan. 1988. *Aristotle: The Desire to Understand*. Cambridge: Cambridge University Press.

Pears, David. 1978. "Aristotle's Analysis of Courage," *Midwest Studies in Philosophy,* 3: .

Roberts, Robert. 1984. "Will Power and the Virtues," *Philosophical Review*, 93: 227–247.

Rorty, Amelie. 1983. "Akratic Believers," *American Philosophical Quarterly*, 20: 175–8.

Scanlon, T. M. 1999. *What We Owe Each Other*. Cambridge: Harvard University Press.

Velleman, David. 1989. *Practical Reflection*. Cambridge: Harvard University Press.

Wiggins, David. 1987. *Needs, Values, Truth*. Oxford: Blackwell.

Williams, S. 1990. "Belief, Desire and the Praxis of Reasoning," *Proceedings of the Aristotelian Society*, 90: 119–42.

Zagzebski, L. 1996. *Virtues of the Mind*. Cambridge: Cambridge University Press.

12

THE VIRTUE OF KNOWLEDGE

Keith Lehrer

The connection between intellectual virtue, justified acceptance, and knowledge requires elucidation. The coherence theory of knowledge that I have developed has a very natural place for the role of intellectual virtue. Moreover, I think I can provide some clarification of the nature of intellectual virtue and resolve some tension between subjective and objective, internalist and externalist, as well as motivational and reliablist approaches to the subject of virtue by considering the role of intellectual virtue in knowledge.[1] That is my objective in this essay. I shall begin with a brief account of the conception of knowledge I have developed. The target of my account of knowledge is what I have in a recent essay called *discursive knowledge*.[2] It is the sort of knowledge a subject can use as a premise in reasoning to confirm some conclusions and reject others. I do not assume that my target conception of knowledge, *discursive knowledge*, is the only one that is important in an overall theory of human cognition. It is, however, a human intellectual product that plays a central role in human science and society.

I am not concerned with information that animals and children possess that they are unable to use in reasoning, though it influences how they respond to the world around them. To avoid the danger of verbal dispute, I do not object to calling such information *primitive knowledge*, but it is not the subject matter of this essay.

I also do not claim that my account of this conception of knowledge is superior to others. This account of knowledge is, however, a useful one for explicating the relationship between intellectual virtue and knowledge. My methodology in philosophy, like my theory of knowledge, is coherentist. It affirms that, rather than asking what is basic, one should explicate the relationships between important conceptions, such as virtue, justification, and knowledge without burdening the discussion with argumentation concerning what is primitive. It is proven, but often forgotten, that whatever you take as primitive in one axiom system can be taken as defined in another system yielding the same theorems and having the same content. We have to start somewhere, of course, but what matters is the elucidation achieved that may be increased by considering how, starting with something else, we may arrive at the

same result. We have no ultimate foundations in philosophy, only usefully chosen beginnings in the loop of explication. I start with knowledge and turn to virtue, but it will become clear that I might have started with virtue instead.

Knowledge and Virtue

Knowledge is undefeated justified acceptance, justified acceptance that cannot be refuted by pointing out any error in what the person accepts, or so I argue. Acceptance that p is a state, a functional state, aimed at accepting p if and only if it is true that p. This truth-directed goal is one that one may pursue in a variety of ways. Acceptance becomes justified when truth is pursued in the right way or in an intellectually virtuous way. People may pursue truth in the wrong way or in an intellectually irresponsible manner. They may reason fallaciously, for example, and accept what they do on the basis of such reasoning. That would not justify them in what they accept even when, as luck would have it, they achieve the goal of accepting that p exactly when it is true that p, just because they accept p and it is true that p. Or suppose that they accept that p because someone has told them p whom they consider to be an authority when they have the clearest evidence of the unreliability of their source of information. There are many wrong or intellectually irresponsible ways to accept what one does even with the motivation of accepting something just in case it is true. Being well motivated is not the same thing as pursuing one's objectives in the right way or in an intellectually virtuous manner.

So what is it to pursue the truth objective in the right way? One answer, a familiar one, is that to pursue the truth in the right way is to pursue it in a way that is reliably successful. This yields one account of virtue, an objective or consequentialist one defined in terms of the successful attainment of the intellectual end of truth. The virtue of the acceptance is to succeed in grasping truth as the virtue of the eye is to see. But I do not think that the right way or the intellectually virtuous way to pursue truth can be equated with reliable success in obtaining truth. One may pursue the truth objective in the right way and in an intellectually virtuous manner in circumstances in which one is invincibly deceived by powerful deceivers, either human or demonic, and if the deception is broad enough in scope, one will not be reliably successful even proceeding in the right way and in an intellectually virtuous manner.

We do, of course, accept that when we pursue truth in the right way or in an intellectually virtuous manner, we do so in a way that is reliably successful, but this accepted connection between the right way or intellectually virtuous manner and the reliably successful way need not hold for us to be justified in what we accept. It must, however, hold for us to be irrefutably justified, but that is a matter that we shall consider later.

So what is the right way or intellectually virtuous manner to pursue truth? We must pursue the goal in a way that is worthy of our trust, which is the same as pursuing it in a way that manifests the intellectual virtues. The principal among these virtues are those of recognizing objections to the candidate acceptance and either

meeting them or falling short of being justified. Justified acceptance is, however, a more restrictive notion than reasonable acceptance. One may, I suggest, reasonably accept things when one cannot meet objections to what one accepts. Reasonably accepting something in this way is not sufficient for being justified in what one accepts (not even prima facie justified) in a way that converts to knowledge if undefeated by errors in what the person accepts. As Austin once remarked, "know," like "can," is an all-in word, and this implies that the kind of justification required for knowledge must be one in which the intellectual virtues have the capacity to meet all objections. One is reasonable to accept many things that one is not justified in accepting in a way that converts to knowledge because one can reasonably accept many things while lacking the capacity to meet all objections to them.

Of course, one must also be ready to change what one accepts when acknowledging objections. Readiness to change when it is appropriate is a virtue. However, such changes must themselves be made in an intellectually virtuous manner. One must change what one accepts in a manner that is worthy of one's trust and that makes one worthy of one's own trust. Moreover, one may also change the way one changes what one accepts and must manifest intellectual virtue in the way one changes the way in which one changes what one accepts. We may avoid a regress by requiring simply that a person must manifest intellectual virtue in how one changes and, correspondingly, be worthy of one's own trust in how one changes. Synchronic virtue at a time points beyond that moment in a present disposition or habit to diachronic virtue over time. That is often noticed. What is less often noticed but equally important is that virtue reaches dynamically beyond the present dispositions to the improvement and correction of them. The dynamics of virtue thrust us into a future of change without a fixed point in the present.

Virtue and Success

What is the connection between proceeding in an intellectually virtuous manner with the purpose of obtaining truth and avoiding error and success in achieving the purpose? How is intellectual virtue connected with intellectual success? The answer is to be found in a theory of justification and knowledge. Let me put the matter in the first person. Suppose I proceed in a completely virtuous way in my acceptance of some claim. I consider all objections to what I accept. I meet all objections. I consider and meet the objections in an intellectually virtuous manner. Nevertheless, through no fault of my own, I may fall into error. I may be deceived. I may, though circumspect, make some mistake that I cannot discern. At the present moment in time with the intellectual resources I have, I may be faultless. I am, moreover, reasonable in what I accept, having followed the principles of reason the best way that I can. Again, I may err, nonetheless, when invincible deception is used against me, for example. I may be faultless but I am not invincible.

So, I arrive at justified acceptance as a consequence of intellectual virtue. Notice that such justified acceptance is not simply a matter of accepting what I cannot help accepting. It would be a mistake to say that a person is justified in accepting any-

thing he cannot help accepting. He may lack virtue because he did not trouble himself to develop any and, as a result, at the present moment cannot help accepting things in an intellectually irresponsible way. There is no successful argument from incapacity to be virtuous to the conclusion that one is, nonetheless, rational or justified in what one accepts. Irrationality and irresponsibility may be ineluctable, but that does not make one rational or responsible. One is not justified in accepting something simply because one cannot help accepting it, for one can lack virtue and fail to accept what one does in a virtuous way even though one cannot help being intellectually irresponsible.

Intellectual virtue may suffice for justified acceptance, but such justified acceptance might not convert into knowledge. The reason is simply that intellectually virtuous acceptance of something, indeed, even completely virtuous acceptance is logically compatible with error. The human intellect and the most virtuous use of it is fallible in striving to obtain truth and avoid error in what one accepts. So a person can accept something in the right way, in an intellectually virtuous manner and in a way that is worthy of trust, and still fail to obtain truth and avoid error.

Moreover, the right way to proceed, the intellectually virtuous way, and the way that makes you worthy of your trust may be ways of accepting things that not only fail in the single instance, but also fail to be reliably successful in general. Whether we are successful will depend not only on the way and manner in which we proceed, but on the local and general circumstances in which we proceed. For intellectual virtue to succeed in fulfilling its purpose of obtaining truth and avoiding error in a reliably successful way, there must be a match between the way of virtue and the way of the world. Of that, there is no necessity. Intellectual virtue is, of necessity, its own reward, but the reward of success in attaining its purpose is mere contingency.

Should we say that a person who proceeds to accept what she does in an intellectually virtuous manner should be satisfied even when confronted with circumstances in which she most frequently accepts what is false? She might obtain some satisfaction from her intellectual virtue, from proceeding in a way that makes her worthy of her own trust, but that is a hollow satisfaction if she generally fails to fulfill the very purpose at which she aims in her intellectual virtue. To be sure, she is virtuous and deserves the satisfaction of virtue, but when virtue is tied to a goal and the rope breaks, the fact that she has done all that she virtuously could to attach herself to the goal leaves her unconnected to the goal of her virtue. The sweetness of such satisfactions of virtue is laced with the bitterness of failure in the very purpose of it.

The Way of Virtue and the Way of Truth

Must we concede, therefore, that when intellectual virtue reliably succeeds in obtaining truth and avoiding error, when trustworthiness succeeds in obtaining truth, this is just a matter of luck? Must we concede that we have no way of determining whether intellectual virtue and trustworthiness yield reliability and success in achieving their purpose? Must we admit that we are ignorant about whether the

way of virtue matches the way of truth? It might seem so, for when we reflect on examples, however fantastical and unrealistic, of universal and invincible deception, we understand that intellectual virtue and trustworthiness would be powerless to expose such deception. So how can we determine that we are not so deceived? And, if we cannot determine that we are not, how can we know that intellectual virtue and trustworthiness ever achieve their purpose? How can we know that the way of virtue is the way of truth? Are we not locked within the internal chamber of the mind with no exit to the external truth beyond?

I have myself thought this was so, but suddenly the error of such reflection manifested itself when I asked myself the simple question — Do I accept that the way of virtue is the way of truth? The question is no sooner posed than answered, for I do accept that the way of virtue matches the way of truth. But is my acceptance just a dogmatic surd, something that I accept without any explanation or justification of it? It must have an explanation and justification. For as I reflect on the matter, I am convinced, not only that I accept that there is the match but that I know that there is. I reflect on what I accept about the external world, simple things, of a table before me, for example, and I am convinced that I do not merely accept that these things exist, but that I am justified in accepting them and know that they exist as well. I know that the way of virtue matches the way of truth, so that my virtuous and trustworthy way of accepting what I do about the objects of the external world reaches the truth. Intellectual virtue and trustworthiness succeed. They achieve the goal of general reliability in accepting what is true and avoiding accepting what is false. But how can I know this?

The answer lies in the account of knowledge offered above. Intellectual virtue suffices for the justification of acceptance. One objection to what I accept is that the way in which I have accepted it is not intellectually virtuous or trustworthy. I reply by meeting it with the answer that I am intellectually virtuous and trustworthy in what I accept as I can tell when I reflect upon the way I proceeded. But then comes the second objection, the more difficult one. It is that, though I am intellectually virtuous and trustworthy in what I accept, my virtue and trustworthiness are unreliable in obtaining truth and avoiding error. How am I to answer this objection? I appeal again to the ways of intellectual virtue and trustworthiness and answer that intellectual virtue and trustworthiness are reliably successful. The answer is a circumspect and cautious one based on experience that has confirmed that the way of virtue matches the way of truth.

Vicious Circles and Virtuous Explanation

Am I arguing in a circle? I am. Is that a vice? It depends on the purpose of it. If I attempt to prove something to another and somewhere assume as a premise the very conclusion I seek to establish to her satisfaction, then I have begged the question, the argument is judged a *petitio*, and the conclusion remains unproven. However, this is more a matter of rhetoric than logic, and the rules of rhetoric depend on the pur-

poses they serve, most notably, to settle disputes in a fair manner. Now it is clearly unfair to assume the very conclusion that is under dispute, so a question begging argument, a *petitio*, must be rejected by the just rules of rhetoric. There are, however, other purposes of reasoning, for example, those of explanation, where the purpose is to explain as much as possible and to leave as little unexplained as one must.

Consider the attempt, now, to explain why what we accept is justified, when it is. Suppose we have a theory of justification to explain why we are justified in accepting what we do when we are justified. What should we answer to the question of whether we are justified in accepting the theory of justification itself? Notice that we accept the theory itself. Thus, if the theory explains why we are justified in accepting what we do, then if we are justified in accepting the theory, the theory must explain why we are justified in accepting it. We might, of course, just affirm that our acceptance of the theory is justified without any explanation of why. But that would leave something unexplained that should be explained by a theory that explains why we are justified in accepting everything we are justified in accepting. A theory of justification either explains why we are justified in accepting the theory itself or the theory remains incomplete. Thus, a complete theory of justification that we are justified in accepting must explain why we are justified in accepting it. The way in which the theory explains why we are justified in accepting it, though it involves a justificatory loop of the theory back on to itself, fulfills the purposes of explanation. We cannot use the theory in a dispute with another to prove that we are justified in accepting what we do. That would be contrary to the rules of rhetoric. But we can and should use the theory to explain why we are justified in accepting it among other things. That is consonant with rules and purposes of explanation.

Now let us consider the application of this point to the issue of intellectual virtue and reliable success in obtaining truth and avoiding error. The objective of intellectual virtue is this kind of success. The aim of virtue is that virtue should succeed and that the way of virtue should be the way of truth. I have argued elsewhere that there is a virtuous loop arising from intellectual virtue and trustworthiness. It is this. Assume I am trustworthy in what I accept. Then, for any target acceptance, acceptance that p where p is the target, I may conclude that I am trustworthy in the target acceptance, acceptance that p. This is not a deduction. I cannot deduce from my general trustworthiness in what I accept, that I am trustworthy in accepting some target, p, because trustworthiness is simply a disposition to be trustworthy in particular cases. It is fallible and does not logically guarantee success. Nevertheless, the reasoning, if not deductive, is, nevertheless, cogent in the way in which an argument from a disposition to the successful exercise of it is cogent.

Acceptance and Trustworthiness

So, assuming that I am trustworthy in what I accept, I can conclude that I am trustworthy in a target acceptance, an acceptance of some specific p. Now the question arises whether I was trustworthy in accepting the initial assumption, that I am trustworthy in what I accept. I can say this. I accept that I am trustworthy in what I ac-

cept. This target acceptance is trustworthy by the general argument form proceeding from my general trustworthiness in what I accept to the target acceptance of my trustworthiness in what I accept. This may look suspicious, and it is important to remember that an unsound argument proves nothing. The premise that I am trustworthy in what I accept must be true for the argument to succeed. However, what the argument shows is that given the truth of the premise that I am trustworthy in what I accept, my trustworthiness in what I accept will explain not only why I am trustworthy in accepting other target acceptances but also, and in exactly the same way, why I am trustworthy in accepting that I am trustworthy in what I accept. Schematically, the argument is as follows:

(1) I am trustworthy in what I accept.
(2) I accept that p.
(3) I am trustworthy in accepting that p. [From (1) and (2)]

Let p = I am trustworthy in what I accept.

(4) I am trustworthy in accepting that I am trustworthy in what I accept. [From (3) and (=)]

My trustworthiness in what I accept explains why I am trustworthy in accepting that I am trustworthy. I must be trustworthy in what I accept in order for the argument to succeed in explaining why I am trustworthy in accepting that I am trustworthy. But then it does succeed.

It is important to notice that the argument is Janus faced and can be extended both forward and backward. We can extend the argument forward to the conclusion that I am reasonable in what I accept, for I am reasonable to accept something I am trustworthy in accepting to achieve a fundamental objective of reason, namely, to accept what is true and avoid accepting what is false. So, I am reasonable to accept what I do when I am trustworthy in accepting it, and my reasonableness is explained by my trustworthiness in accepting it.

Again, schematically, the argument is as follows:

(1R) I am trustworthy in what I accept.
(2R) I am reasonable in what I accept. [From (1R)]
(3R) I accept that p.
(4R) I am reasonable in accepting that p. [From (2R) and (3R)]

Let p = I am trustworthy in what I accept.

(5R) I am reasonable in accepting that I am trustworthy in what I accept. [From (4R) and (=)]

Thus, the reasonableness of my accepting that I am trustworthy in what I accept is explained by my trustworthiness in what I accept.

We can extend the argument backward a bit by inquiring why I am trustworthy in what I accept. Now here I might appeal to the trustworthiness of how I proceed, but the question of why I am trustworthy in proceeding in those ways and other questions about why I am trustworthy in doing this or that, aiming at this or that end, using this or that means to attain the end, and so forth, can only come to rest in the principle that I am trustworthy. It is the trustworthiness of the self that is the ultimate principle of explanation. Moreover, it has the merit mentioned above. I accept that I am trustworthy, and I conclude that I am trustworthy in what I accept for the purpose of obtaining truth and avoiding error. My trustworthiness explains why I am trustworthy in what I accept, and, indeed, reasonable in what I accept for this purpose.

Schematically again, the argument is as follows:

(1S) I am trustworthy.
(2S) I am trustworthy in what I accept. [From (1S)]
(3S) I accept that I am trustworthy.
(4S) I am trustworthy in accepting that I am trustworthy. [From (2S) and (3S)]
(5S) I am reasonable in accepting that I am trustworthy. [From (4S)]

There is a loop in the explanation, but this is the virtuous loop of explanation rather than the vicious circle of proof.

Let us reconsider this briefly in terms of intellectual virtue. Suppose that I proceed the way I should, manifesting a disposition of intellectual virtue in what I accept. (As well as in how I reason, and, for that matter, in what I prefer to accept as well.) Now the exercise of intellectual virtue will lead me to accept that I am virtuous in what I accept and, moreover, in accepting that I am virtuous in accepting that I am virtuous in what I accept. My disposition of intellectual virtue will explain why I am intellectually virtuous in accepting what I do, including the target acceptance of my intellectual virtue. The argument will again lead to the conclusion that I am reasonable in what I accept, and the reasonableness will be explained in terms of my intellectual virtue, my disposition to be that way. Similarly, we may appeal to my general virtue to explain my intellectual virtue. These explanatory arguments are not deductive. I can be virtuous and fail to be intellectually virtuous just as I can be intellectually virtuous in general and fail in the particular instance to be intellectually virtuous. Reasoning from dispositions to the successful exercise of them is subject to the hazards of our fallibility.

The reasonableness and, finally, the justification for accepting what we do, including that we are intellectually virtuous and trustworthy, is a consequence of our being intellectually virtuous and trustworthy. The reward of intellectual virtue and trustworthiness is that it can vouch for itself in the life of reason and the explanation of justification in a way that is intellectually virtuous and trustworthy. Of course, the explanation is capable of confirmation and requires it. The loop must widen to include the confirmation of what we accept in support of the acceptance of our virtue and trustworthiness. We do not become intellectually virtuous by accepting that we are. Virtue is the result of works, in this case of what we accept and what we reject, which confirms that we have the disposition of intellectual virtue and trustworthi-

ness. There is, of course, no paradox in the idea that a general principle that explains particular instances, such as a law of physics, is confirmed by the instances it explains. The trustworthiness, virtue, and reasonableness of the instances confirm the presence of the disposition of intellectual virtue, trustworthiness, and reasonableness when these are present.

Virtue and Justification

Intellectual virtue can provide us with the means to answer all objections to some, though not all, target acceptances in a reasonable way. When intellectual virtue is sufficient to answer all objections, even if not exercised to do so, we are justified in a way that will convert to knowledge provided that the justification cannot be refuted or defeated by some error in what the person has accepted to obtain the justification. This, of course, brings us to problems of global deception made famous by Descartes as well as to more local problems of deception made famous by Gettier. Deception and error can refute and defeat our reasonable and justified acceptance. This fact draws us out of the virtuous loop of the explanation of our reasonable and justified acceptance to the fundamental question. Are we reasonable and justified in accepting that such virtue and trustworthiness hit their mark? Are we reasonable and justified in accepting that virtue and trustworthiness achieve their purpose of obtaining truth and avoiding error? Can we know that our intellectual virtue and trustworthiness achieve this purpose?

The affirmative answers to these questions result from an expansion of the explanatory loop. Let us consider the questions in terms of intellectual virtue. Suppose I have a virtuous disposition to fulfill my purpose of accepting what is true and avoiding accepting what is false. As an excercise of that disposition, I accept many things about the world, about what is true and what is false. One thing I accept is that a virtuous disposition to fulfill my purpose is rewarded with reliable success in obtaining truth and avoiding error. Moreover, I am intellectually virtuous in accepting this. I note that intellectual virtue is fallible, and there is no logical guarantee available for the reliable success of intellectual virtue. But the lack of a logical guarantee and the fact of our fallibility does not in any way preclude the only sort of assurance available, namely, that of an intellectually virtuous disposition. We seek to be intellectually virtuous in a way that will lead to reliable success. That is the primary purpose of intellectual virtue, and if fails to achieve the purpose, then virtue does not receive the reward of its purpose. So, as we proceed, we assume that intellectual virtue leads to success, and we would lose interest in it otherwise.

Put in the simplest terms, we accept that intellectual virtue is reliably truth connected. What we accept confirms that it is. For we find that intellectual virtue, though fallible, is reliably connected with truth, and we accept the finding. Now assume that this is true, that intellectual virtue and trustworthiness are reliably truth connected. Assume as a premise for explanation that intellectual virtue is reliably truth connected. Then we can explain why it is reasonable for us to accept that intellectual virtue is reliably truth connected. It is because our acceptance of this, of the

successful reliability of the truth connection, is explained by the successful reliability of the truth connection.

Here again is an argument from a general disposition to a particular instance. The disposition to be intellectually virtuous in what I accept is reliably truth connected. So when I accept some target acceptance, that p, I conclude, first, by the previous argument about intellectual virtue, that I am intellectually virtuous in accepting that p, and, second, by the present argument that my intellectual virtue in accepting that p is a reliably truth-connected kind of acceptance. The argument is, of course, only sound if its premises are true. The premise that intellectual virtue is reliably truth connected must be true for the argument to be sound. Furthermore, the argument for the conclusion that I am intellectually virtuous in accepting that p in a way that is reliably truth connected is, though cogent, not a deductive demonstration. The explanation for the successful reliability of my intellectually virtuous kind of acceptance that p is the successful general reliability of my intellectually virtuous kind of acceptance.

General Reliability and Intellectual Virtue

Is the general reliability of intellectual virtue and trustworthiness sufficient for the kind of truth connection we require to convert justification to knowledge? Unfortunately, it is not. The reason is that our intellectual virtue and trustworthiness may be reliably successful in leading us to accept what is true and avoid accepting what is false but may fail to explain why what we accept is true in the particular instance. Consider a person, Mr. Goodsumer, who is intellectually virtuous in a reliably successful way in the manner in which he adds up numbers. He carefully double checks what he does, summing up and down the column of numbers, but in a particular instance, though he proceeds in the usual trustworthy and reliably successful manner, he makes a mistake, two perhaps, and gets the right answer by luck.

Mr. Goodsumer accepts that his intellectually virtuous and trustworthy way of accepting the answer is successfully truth connected, but in this case, it is not, even though what he accepts as the sum is correct. His way of doing sums is reliably successful—he almost never makes mistakes—but in this instance he made mistakes, though by luck he got the right answer anyway. The problem is that he did not get the correct answer *because* he added the numbers in an intellectually virtuous and trustworthy way. He was not successful because he proceeded in a way that is reliable. Mr. Goodsumer is successful because of luck.

The explanation of why he does not know on our account of knowledge is easy enough to provide. He accepts that he is successful in getting the right answer because he has proceeded in a intellectually virtuous and, hence, reliable way. But this is false. That is not why he was successful in getting the right answer, and the falsity of it suffices to refute and defeat his justification. The only problem is to explain what it means to say that a person is successful in accepting what is true because she has proceeded in a trustworthy and reliably successful manner.

We have explained what it is that makes a person intellectually virtuous and trustworthy in what she accepts. It is to have a disposition of a certain sort aiming at accepting what is true and avoiding accepting what is false. What makes a person who is intellectually virtuous and trustworthy in what she accepts successfully reliable as well? To be successfully reliable is to generally succeed in achieving the goal of accepting what is true and avoid accepting what is false. It is, therefore, to have a high frequency of success in accepting what is true and avoiding accepting what is false, perhaps resulting from an underlying propensity toward that frequency.

What does it mean to say that trustworthiness in what one accepts is successfully connected with truth in what one accepts in a particular case? It cannot mean, as we have noted in the case of Goodsumer, that being trustworthy in what one accepts is generally or reliably successful. It means, instead, that the person is successful in accepting what is true because she accepts what she does in an intellectually virtuous and trustworthy way in the particular case. The explanation of her success in that particular case is her virtuousness and trustworthiness. Perhaps all we can say in addition by way of explanation of what this means is that it is the manifestation of her disposition to be intellectually virtuous and trustworthy and the underlying propensity of her disposition to lead her to accept what is true and avoid accepting what is false that explains why she accepted something true in this instance. Her virtuousness and the reliability of it explains her success in the particular case.

Reliablism and Justification

It may be important to notice that reliablism alone will not suffice for the sort of justification required for knowledge. First of all, reliability may be opaque to the subject. The virtuousness and trustworthiness of the subject involves the positive evaluation of a target acceptance on the basis of her background system, what I have called her *evaluation system*. This evaluation is an exercise of intellectual virtue that provides the transparency needed for justification, what I have called *personal justification*,[3] even if the transparency of such justification does not extend to the undefeated or irrefutable feature that justification must possess for conversion to knowledge.

Another equally important reason why reliablism will not suffice is that global reliability might be irrelevant locally. Consider again the case of Mr. Goodsumer who sums reliably but not in the particular case. Virtuous and trustworthy acceptance must be connected with truth in the particular instance. What is the required connection? It is explanation. Being virtuous and trustworthy in the particular instance of acceptance *explains* why one succeeds in obtaining truth. The way of virtue is connected with the way of truth in this particular instance. I must succeed in accepting what is true in the particular case *because* I am intellectually virtuous and trustworthy in what I accept *in this case*.

Thus, it is not enough to point to some general success to explain why I succeed, to any general probability, however objective and lawlike. The explanation of success in obtaining truth in what one accepts in a particular case depends on the features of the particular case. A scientific explanation of why we succeed in obtaining

truth in what we accept in a particular instance may admit of the appeal to an objective probability or scientific law, but this is not essential for knowledge. We know ordinary simple things, that we see a table, that we have done a simple sum correctly, that we are speaking with another, before we know science. Indeed, we even know that we know our virtuousness and trustworthiness in what we accept is connected with truth in the particular instance before we can provide any scientific explanation of why this is so. We often know that x occurs because y does and that the occurrence of y explains the occurrence of x when we cannot provide any scientific explanation of why it is so. Science may demand such an explanation, and it is much to be desired, but epistemology can proceed without it.

Intellectual Virtue and the Truth Connection

Now we may, having qualified the argument concerning the truth connection, apply that argument to the special case in which the acceptance of p is the acceptance that my intellectual virtue in what I accept is reliably truth connected. The reliability of that connection explains why my acceptance of the reliability of that connection is reliably truth connected. Moreover, and most crucially for the issue of reasonableness, justification, and knowledge of that connection, we can reach felicitous conclusions. I am reasonable in accepting that virtuous acceptance is reliably connected with truth. I aim at reliably truth-connected virtue in what I accept, and, assuming I achieve it, I am reasonable to think I do. This is not an argument to the effect that if something is true, then it is reasonable for me to think it is true. It is an argument to the effect that because intellectual virtue succeeds in its purpose of being reliably truth connected, therefore, when intellectual virtue leads us to accept something, we may cogently, though not deductively, conclude that it is reasonable to accept it.

It will be useful, once again, to have the argument schematically presented. The argument is the following:

(1E) What I accept in an intellectually virtuous (trustworthy) way is reliably truth connected.
(2E) I accept that p in an intellectually virtuous (trustworthy) way.
(3E) It is true that p. [From (1E) and (2E)]

Let p = What I accept in an intellectually virtuous (trustworthy) way is reliably truth connected.

(4E) It is true that what I accept in an intellectually virtuous (trustworthy) way is reliably truth connected. [From (2E), (3E), and (=)]
(5E) It is reasonable for me to accept that what I accept in an intellectually virtuous (trustworthy) way is reliably truth connected. [From (2E) and (=)]

It is important to note and to concede that crucial steps in this argument and in all the schematically presented arguments depend on inference from a general disposition or general capacity to be trustworthy or reliably truth connected to conclusions about the trustworthiness and truth connectedness in particular instances. These inferences are inductive or ampliative rather than deductive or demonstative, for we are not always trustworthy in particular cases though we are trustworthy in general. So, what we accept in a virtuous (trustworthy) way is not always true in particular cases though our virtuous (trustworthy) acceptance is, in general, reliably truth connected.

Nevertheless, the inference from a general capacity to the successful exercise of it is a cogent and reasonable one even if it is fallible and defeasible. The success of the argument depends on whether our success in obtaining truth in the particular case, including the case of the reliable truth connectedness of intellectual virtue, is explained by our intellectual virtue. Truth, reasonableness, justification, and knowledge are tied up, down, and together in a loop explaining how the way of virtue matches the way of truth.

Knowing We Are Not Deceived

So what about deception? I accept that I am not deceived, which may, if I am intellectually virtuous, suffice for meeting the objection. What about the possibility of deception? The possibility of deception, which is closely connected with our fallibility, argues for the conclusion that whenever we accept that we are not deceived, we could have been deceived, indeed, invincibly deceived. Therefore, accepting we are not deceived is something we would accept even if it were false. That is the nature of deception. I am not deceived, however, and my intellectually virtuous acceptance is reliably truth connected. That is why I know that I am not deceived.[4]

All this concerns what I accept in an intellectually virtuous way. If someone demands that I should not appeal to what I accept to answer a skeptic, then I would have to put my hand over my mouth to comply with his demand. But why should I comply? What I accept is accepted in a way that is intellectually virtuous; it represents my best efforts to accept truth and avoid error; and why should I silence my voice of virtue to accommodate an unjust demand?

But, one might object, if you are in error in what you accept, then your justification based on your intellectually virtuous acceptances, your meeting of objections, however virtuous, is useless for that attainment of knowledge. False premises are the instruments of ignorance, not knowledge. To which I must reply, in candor, that if I am in error in what I accept, then, yes, I am ignorant. However, I am not ignorant because I am not in error, at least, in enough of the things I accept. As a result, justification of acceptance in terms of enough truth in what I accept will often prove irrefutable and undefeated by errors in my acceptance. To be sure, intellectual virtue must combine with intellectual fortune so that the way of virtue is also the way of truth. I am not, however, ignorant of my good fortune, but accept it with gratitude as an exercise of virtue.

I am justified in accepting that intellectual virtue and trustworthiness are reliably connected with truth, reliably successful in accepting what is true and avoiding accepting what is false. Such justification converts to knowledge because there is no demon, no grand deceptions, though some local deceptions and errors cause to me to fall ignorant in some instances. I often know what I accept is true because intellectually virtuous and trustworthy acceptance is reliably truth connected. As a result, I know that intellectual virtue and trustworthiness are reliably truth connected, and I know that I know this, even though I am fallible in what I accept and use to meet the objections of the internal and external skeptic. You do not have to be infallible to be reliable, or to know, or to know that you know, about many things. The most important thing is that intellectual virtue succeeds in its purpose of obtaining truth and avoiding error in the way we accept that it does. Such success explains why we have obtained truth in general and why knowledge is transparent to us in the particular instances of it. It is good to know it.

Notes

1. I am much indebted to the work of Ernest Sosa, Alvin Goldman, Linda Zagzebski, John Greco and, more remotely, Thomas Reid. Their works are cited along with the work of many others in "Recent Work on Virtue Epistemology," Guy Axtell, ed. *American Philosophical Quarterly,* 34, 1 (1997): 1–26. I refer the reader to this work for references rather than repeating them here except for mentioning my special indebtedness to Ernest Sosa, *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991); Alvin Goldman, *Liasons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, Mass.: MIT Press, 1992); Linda Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press, 1996); and John Greco, *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry* (Cambridge: Cambridge University Press, forthcoming.) There is some overlap between later sections of this essay and Chapter 9 of *Theory of Knowledge, Second Edition* (Boulder: Westview Press, forthcoming). I am also indebted to David Truncellito for noting the importance of the notion of accepting something in the right way, which he explored in his doctoral dissertation that I directed.
2. Discursive Knowledge, forthcoming in *Philosophy and Phenomenological Research*. This conception of knowledge was the target of my earlier analysis of knowledge in *Theory of Knowledge* (Boulder: Westview Press, 1990), and most recently in "Justification, Coherence and Knowledge," *Erkenntnis* (1999): 1–16.
3. In *Theory of Knowledge* (Boulder: Westview Press, 1990).
4. Cf. "How to Defeat Opposition to Moore," *Philosophical Perspectives 19: Epistemology,* James Tomberlin, ed. (Atascadero, Calif.: Ridgeview Press, Fall 1999).

13

THE FOUNDATIONAL ROLE OF EPISTEMOLOGY IN A GENERAL THEORY OF RATIONALITY

Richard Foley

An unfortunate methodological assumption of much recent epistemology is that the properties that make a belief rational are by definition such that when a true belief has these properties, it is a good candidate for knowledge, with some other condition added to handle Gettier-style counterexamples. This assumption has the effect of detaching the theory of rational belief from a general theory of rationality and placing it instead in service to the theory of knowledge. If it is stipulated that the properties that make a belief rational must also be properties that turn true belief into a good candidate for knowledge, then an account of rational belief can be regarded as adequate only if it contributes to a successful account of knowledge. The theory of rational belief is thus divorced from everyday assessments of the rationality of decisions, plans, actions, and strategies, and it is even divorced from everyday assessments of the rationality of opinions, which tend to focus on whether individuals have been responsible in forming their opinions rather than on whether they have satisfied the prerequisites of knowledge.

The remedy is for epistemologists, at least at the beginning of their enterprise, to be wary of the assumption that knowledge can be adequately understood in terms of rational true belief plus some fillip to handle Gettier problems. By the end of the enterprise, after accounts of rational belief and knowledge have been independently developed, interesting connections between the two may have emerged, but it ought not simply be assumed from the start that there is a simple, necessary tie between them. Trial separations are often liberating for both parties, and this one is no exception. Relaxing the tie between the two frees the theory of knowledge from overly intellectual conceptions of knowledge, thus smoothing the way for treatments that adequately recognize that people are not in a position to provide a justification for much of what they know, and it simultaneously creates a space for the theory of rational belief to be embedded in a general theory of rationality.

My focus here will be on the latter benefit. The notion of rational belief ought not be cordoned off from other notions of rationality, as if the conditions that make a belief rational have little to do with the conditions that make a decision, action, strat-

egy, or plan rational. The way we understand the rationality of beliefs ought to be of a piece with the way we understand the rationality of other phenomena.

A first step toward a well-integrated theory of rationality is to recognize that rationality is concerned with the effective pursuit of valuable ends, what I shall call "goals." However, it is too stringent to insist that rationality requires one actually to succeed in satisfying one's goals. Consider plans, for example. Rational plans sometimes turn out badly. Indeed, a plan can be rational even if one's ends are unlikely to be achieved by the plan; it may be that no one could be reasonably expected to see that the plan was likely to have these consequences. Considerations such as these suggest a general schema of rationality:

> A plan (decision, action, strategy, belief, etc.) is rational for an individual if it is rational for the individual to believe that it would satisfy his or her goals.

The rationality in question here is abstract as opposed to concrete, in the sense that uninstantiated plans, decisions, actions, etc., can meet the conditions of the schema. Even if I have not adopted a plan P, it may nonetheless be rational for me to believe that P would have satisfied my goals. If so, the uninstantiated, unchosen P satisfies the above schema. Similarly, a plan can conform to the schema even if it is adopted for the wrong reasons. Suppose it is rational for me to believe that plan Q would satisfy my ends, but I choose Q only because my Ouija board recommends it. Plan Q is rational in the relevant abstract sense (it is the plan which it is rational for me to prefer) even if its concrete realization (the process by which I came to choose Q) leaves much to be desired.

Because there is more than one sense of rationality, the schema does not purport to express a necessary condition of rationality, only a sufficient condition. Assessments of rationality can be made from various perspectives, and corresponding to these perspectives are different senses of rationality and reasons.[1] If a plan would in fact satisfy my goals, then in one sense of reason, a fully objective sense, there is a reason for me to adopt the plan, even if I am not in a position to recognize that it would produce these benefits. At the other extreme, if I am genuinely convinced that a plan would satisfy my goals, then in one sense of reason, this time a fully subjective sense, I have a reason to adopt the plan, even if there is little or no evidence to indicate that it will in fact achieve my goals. After all, doing anything else would be directly counter to what I myself think best, which is the height of subjective irrationality. Nevertheless, although there are occasions on which we find it natural to make assessments of reasons and rationality from a fully objective perspective (where all that matters is whether the plan, decision, etc., will in fact achieve the desired ends) and other occasions on which we find it natural to make such assessments from a fully subjective perspective (where all that matters is whether the individual believes that the plan, decision, etc., will achieve the desired ends), it is more common in everyday life, in the law, and in other contexts for us to be interested in what it is reasonable for the individual to think of the plan, decision, etc. These are the assessments that the above schema represents.

An obvious drawback of the schema is that it includes a reference to the notion

of rational belief. It thus leaves us within the circle of notions we wish to understand. For the moment, however, I will set aside this problem, because there are other questions about the schema that also need to be addressed. For instance, for a plan, decision, action, strategy, belief, etc., to be rational, must it be rational to believe that it does a better job of achieving one's goals than any of the alternatives, or might something less than the very best do? As I will be using the terms, "reasonableness" admits of degrees whereas "rationality" does not. In particular, reasonableness varies with the strengths of one's reasons, and the rational is that which is sufficiently reasonable. This usage has the welcome consequence of leaving open the possibility that several options might be rational for an individual even though there are reasons to prefer some of the options over others. Thus, the refined schema of rationality is:

> A plan (decision, action, strategy, belief, etc.) is rational for an individual if it is rational for the individual to believe that it would acceptably contribute to his or her goals.

To say that a plan would "acceptably contribute" to one's goals is to say its estimated desirability is sufficiently high, where estimated desirability is a function of what it is rational to believe about the probable effectiveness of the plan in promoting one's goals and the relative value of these goals.[2] Contextual factors are also relevant. The estimated desirability of the plan must be sufficiently high given the context, where the context is determined by the relative desirability of the other available options and their relative accessibility. The fewer alternatives there are with greater estimated desirabilities, the more likely it is that the plan in question is rational. Moreover, if these alternatives are only marginally superior or cannot be readily implemented, it is all the more likely that the plan is rational. It will be rational because it is good enough, given the context.

The schema can be further refined. When assessing the rationality of a decision, strategy, etc., we can take into consideration all of the individual's goals or only a subset of them. Often, we do the former. We assess what is rational, all things considered. Sometimes, however, our focus is more narrow. We are interested in how effectively someone is pursuing a specific type of goal. For example, we may want to evaluate a plan exclusively with respect to goals that concern an individual's economic well-being. If we judge that the plan is an effective way of promoting this subset of goals, we can say that the plan is rational in an economic sense for the individual. We can say this even if, with respect to all the person's goals, both economic and non-economic, it is not rational to adopt the plan. Thus, the general schema of rationality needs to be qualified to reflect that there are different kinds of rationality corresponding to various kinds of goals:

> A plan (decision, action, strategy, belief, etc.) is rational in sense X for an individual if it is rational for the individual to believe that it would acceptably contribute to his or her goals of type X.

Distinguishing different types of rationality is especially important when it is the rationality of beliefs that is at issue. In evaluating each other's beliefs, we typically focus attention on intellectual goals rather than pragmatic ones. For example, in assessing whether it is rational for you to believe a hypothesis H, as a rule I am not interested in considering whether believing H would produce economic, psychological, or health benefits for you. More notoriously, in assessing whether it might be rational for you to believe in God, I am unlikely to join Pascal in regarding as relevant the possibility that you might increase your chances of salvation by being a theist. Nevertheless, the above general schema of rationality implies that there is nothing in principle wrong with evaluating beliefs in terms of how well they promote one's non-intellectual goals. This is an important point to which I will return later.

A point of more immediate interest, however, is that in their accounts of epistemically rational belief, epistemologists have traditionally been concerned with not just any intellectual goal, but rather a very specific intellectual goal, that of now having beliefs that are both accurate and comprehensive. This goal, which I shall call "the epistemic goal," is concerned with the accuracy and comprehensiveness of one's beliefs at the current moment, not with their accuracy and comprehensiveness at some future time. To understand why this restriction is important, imagine that one's prospects for having accurate and comprehensive beliefs in a year's time would be enhanced by believing something for which one now lacks adequate evidence. For example, suppose that P involves a much more favorable assessment of my intellectual talents than my evidence warrants, but suppose also that believing P would make me more intellectually confident than I would be otherwise, which would make me a more dedicated inquirer, which in turn would enhance my long-term prospects of having an accurate and comprehensive belief system. Despite these long-term benefits, there is an important sense of rational belief, indeed the very sense that traditionally has been of the most interest to epistemologists, in which it is not rational for me now to believe P. Moreover, the point of the example is not affected by shortening the time period in which the benefits are forthcoming. It would not be rational, in this sense, for me to believe P if we were instead to imagine that believing P would somehow improve my prospects for having accurate and comprehensive beliefs in the next few weeks, or in the next few hours, or even in the next few seconds. The precise way of making this point, in light of the above distinctions, is to say that in such a situation, it is not rational in a purely epistemic sense for me to believe P.

There are competing views about how this purely epistemic sense of rational belief is best explicated. Classical foundationalists have one view. Coherentists have another. Probabilists, reliabilists, modest foundationalists, and virtue theorists have yet other views. For our purposes here, however, it is not important which of these approaches is best, because my primary concern is to illustrate how the notion of epistemically rational belief, whatever precise account one gives of it, plays a crucial role in the general theory of rationality. In particular, I will be arguing that epistemic rationality serves as a theoretical anchor for other notions of rationality.

According to the general schema, a plan, (decision, action, strategy, belief, etc.) is

rational in sense *X* for an individual if it is rational for the individual to believe that the plan (decision, action, strategy, belief, etc.) would do an acceptably good job of promoting his or her goals of type *X*. This schema, which makes use of the notion of rational belief, leaves us within the circle of notions we wish to understand. However, accounts of epistemically rational belief standardly do not, and should not, make use of any other notion of rationality, or any cognate of rationality (warrant, justification, etc.), in their explications of epistemic rationality. For example, classical foundationalists understand epistemic rationality in terms of deductive relations plus direct acquaintance or infallible belief; coherentists understand the notion in terms of belief plus deductive and probabilistic relations; reliabilists understand the notion in terms of the propensity of cognitive processes to generate true opinions; and so on. These accounts make no reference to any other notion of rationality, and thus they provide the above schema with an escape route from circularity. In particular, with an account of epistemically rational belief in hand, the general schema of rationality can be further refined:

> A plan (decision, action, strategy, belief, etc.) is rational in sense *X* for an individual if it is epistemically rational for the individual to believe that it would acceptably contribute to his or her goals of type *X*.

The refined schema illustrates how epistemic rationality can serve as an anchor for other kinds of rationality. Moreover, the schema is perfectly general. It applies to all phenomena (plans, decisions, strategies, etc.) and to all forms of rationality for these phenomena (economic rationality, rationality all things considered, etc.). Most relevant for my present purposes, the rationality of belief is itself an instance of the schema. Even epistemically rational belief is an instance. For example, inserting the epistemic goal into the general schema for "goals of type *X*" results in the following:

> Believing *P* is rational in an epistemic sense if it is epistemically rational for one to believe that believing *P* would acceptably contribute to the epistemic goal of one's now having accurate and comprehensive beliefs.[3]

This instantiation of the general schema is compatible with all the major theories of epistemically rational belief. According to coherentists, it is epistemically rational for one to believe that believing *P* would acceptably contribute to the epistemic goal of one's now having accurate and comprehensive beliefs only when the proposed coherentist conditions are met with respect to the proposition *P*, that is, only when *P* coheres appropriately with one's other beliefs and hence it is epistemically rational to believe that *P* is true. According to reliabilists, it is epistemically rational for one to believe that believing *P* would acceptably contribute to the epistemic goal only when the recommended reliabilist conditions are met with respect to *P* and hence it is epistemically rational to believe that *P* is true; similarly, for classical foundationalists, modest foundationalists, and probabilists.[4]

A potential objection against this schema for epistemically rational belief is that goals are necessarily concerned with future states of affairs, and thus it is not possi-

ble to understand epistemic rationality in terms of one's beliefs contributing to the present-tense goal of now having accurate and comprehensive beliefs.

However, I am using "goal" in a very broad sense, to refer to whatever is of value for an individual. I won't say much about how something comes to have value, because one can employ pretty much whatever account of value one favors in conjunction with the above schemas, as long as the account allows that for most of us most of the time, it is a good thing, all else being equal, to believe P if P is true (the value of comprehensiveness), and likewise a good thing, all else being equal, for P to be true if one believes P (the value of accuracy). I am presupposing, in other words, that for most of us most of the time, having comprehensive and accurate beliefs is prima facie valuable and, hence, a goal.

Because I am equating goals with whatever is of value for an individual, goals can be concerned with current states of affairs as well as future states of affairs. Correspondingly, there can be constitutive means to goals as well as causally effective means to them. If someone has the goal of being in good health, and if good health is a state in which one not just currently lacks disease but also is not prone to disease, then not having high blood pressure is not so much a causal means to the goal of good health but rather part of what constitutes good health. In an analogous manner, now believing P can be a part of what constitutes now having accurate and comprehensive beliefs.

Nevertheless, if anyone thinks that this broad construal of goals is stretching the standard meaning of "goal" too far, it is easy enough to devise alternative terminology. "Goal" in the above schemas can be replaced with "desideratum" or "value," and "contribute" with a locution about what is appropriate or fitting, given this desideratum. The overall schema for understanding the epistemic rationality of a belief P thus becomes: Believing P is rational in an epistemic sense if it is epistemically rational to believe that believing P is appropriate (that is, fitting) insofar as it is a desideratum (that is, a valuable state of affairs) for one now to have accurate and comprehensive beliefs.

Epistemic rationality is defined in terms of a synchronic intellectual goal, but medium- and long-term intellectual goals are also important to us. We want our intellectual skills to develop and our beliefs to improve over time, and thus we seek out strategies that will help us achieve these goals. Temporally extended strategies are as essential in the intellectual realm as they are in the practical realm. It is important, to cite but one example, to have stable, dependable strategies governing the gathering and processing of information.

As important as temporally extended intellectual strategies are, however, they do not eliminate the need for a notion of rational belief that is defined in terms of the synchronic goal of now having accurate and comprehensive beliefs. Even when we are in the midst of trying to improve our epistemic situation, for example, by gathering additional information about a topic, we may wonder what it is rational for us to believe in the interim. Our having reasons to continue inquiries on a topic does not necessarily mean that we have to withhold judgment until we complete these inquiries. We often have good epistemic reasons to believe a proposition even though for safety's sake we also have good reasons to gather additional information. Besides,

whether we have good reasons to gather more information about a topic is itself a function of what it is epistemically rational for us to believe about the consequences of doing so. It may be that taking the time to gather additional information would greatly improve the quality of our beliefs, but it won't be rational for us to do so unless it is epistemically rational for us to believe that the estimated benefits of gathering this information are acceptably high relative to alternative ways of spending our time. So, important as questions of intellectual progress and improvement are, questions of the synchronic rationality of our beliefs have a theoretical priority.

The general schema of rationality implies that the rationality of a plan, decision, action, strategy, belief, etc., is a matter of its being epistemically rational for one to believe that it would do an acceptably good job of satisfying one's ends. Recall, however, that in applying the general schema, we can take into consideration either all of an individual's goals or only a subset of them. This creates a risk of confusion. If we take into consideration only economic goals, for instance, we may judge that it is rational, in an economic sense, for the individual to adopt plan P, but if we take into consideration all of one's goals, both economic and noneconomic, we may well conclude that it is not rational, all things considered, for the individual to adopt P.

These same possibilities for confusion arise, and indeed arise all the more acutely, when the rationality of beliefs is at issue. Beliefs can be assessed in terms of how well they promote the epistemic goal, but as I have already observed, they can also be assessed in terms of how well they promote other goals. For example, inserting "the total constellation of one's goals" into the schema for "goals of type X" yields the following:

> Believing P is rational, all things considered, if it is epistemically rational for one to believe that believing P acceptably contributes to the total constellation of one's goals.

There are two notions of rational belief at work in this characterization. The first is the anchoring notion of epistemically rational belief, defined in terms of the epistemic goal. The second is the derivative notion of rational belief all things considered, defined in terms of the anchoring notion and the complete set of one's goals.

A common complaint against epistemology is that its issues are too rarified to shed much light on the everyday assessments we make, and need to make, of each other's beliefs. Epistemic rationality is concerned with a narrowly circumscribed, synchronic intellectual goal, that of now having accurate and comprehensive beliefs, whereas our everyday assessments are sensitive to the fact that because we have many goals, pragmatic as well as intellectual and long-term as well as short-term, there are sharp limitations on how much time and effort it is feasible to spend pursuing intellectual goals. In contrast to the notion of epistemically rational belief, the notion of rational belief all things considered is characterized in terms of the total constellation of one's goals. So, it is tempting to think that this notion might provide a way of addressing the above complaint. However, I will be arguing that this is not in fact the case. Although there is nothing inherently improper in assessing beliefs in

terms of how well they promote the overall collection of one's goals, in fact we tend not to do so.

At first look, this may seem puzzling. After all, even when it is not epistemically rational to believe that a proposition *P* is true, it can be epistemically rational to believe that the overall consequences (long term as well as short term, and pragmatic as well as intellectual) of believing *P* would be significantly better than those of not believing *P*. Thus, what it is rational to believe, all things considered, can differ from what it is epistemically rational to believe. Nevertheless, conflicts of this sort rarely occur. In all but a few extreme situations, the benefits that might accrue from having beliefs that are not epistemically rational are outweighed by the costs associated with the decline in the overall accuracy of one's belief system.

To be sure, there are conceivable exceptions. If a madman will kill my children unless I come to believe a proposition *P* for which I lack evidence, then I have good reasons to find some way of believing *P*. However, it is not a simple matter to get oneself to believe something for pragmatic reasons. Becoming fully convinced that there is powerful evidence in favor of *P*'s truth is ordinarily sufficient to occasion belief in *P*, whereas becoming fully convinced that there are powerful pragmatic considerations in favor of believing *P* is ordinarily not sufficient to occasion belief, a point which was not lost on Pascal. He was under no illusion that his pragmatic argument for theistic belief would directly convince anyone to believe in God's existence. His strategy, rather, was to convince non-believers to manipulate their situation so that belief would eventually become possible for them. They might do so, he suggested, by surrounding themselves with believers, attending religious services regularly, reading religious literature, listening to religious music, and so on. It is no different with non-theological propositions. If one is to come to believe a proposition *P* for pragmatic reasons, as opposed to evidential reasons, one will usually need to engage in Pascalian manipulations of one's evidential situation, with the aim of altering the situation in such a way that one will ultimately become convinced that there is after all good evidence for *P*'s truth.

However, such manipulations require one to plot against oneself. For example, one may have to find a way to forget or at least to downplay one's current evidence against *P*. Tampering in this way with one's evidential situation will produce what one would now regard as inaccuracies or gaps in one's beliefs system. Moreover, because beliefs cannot be altered in piecemeal fashion, the inaccuracies or gaps will typically extend well beyond *P* itself. Beliefs come and go in clusters. As Peter van Inwagen once observed, one cannot believe that the moon is made of green cheese without also believing all sorts of other, equally astounding claims, for example, that there are (or were) enormously large cows capable of producing prodigious quantities of milk; that there is (or was) an immense cheese processing plant capable of turning this milk into cheese and forming it into a huge sphere; and so on.

Getting oneself to believe a proposition *P* for which one lacks evidence is thus not something one can do in isolation from one's other beliefs. It requires a project that involves altering one's opinions toward a wide number of related propositions as well. From one's current perspective, such a project will be at the expense of the overall accuracy or comprehensiveness of one's belief system, and because effective

decision making normally requires accurate and comprehensive beliefs, this in turn is likely to affect adversely the overall success one has in promoting one's goals. If my children's lives are at stake, costs such as these will be well worth paying, but in less extreme situations, they will not be. The costs associated with the manipulations will be unacceptably high relative to the benefits of the resulting belief.

So, although what it is rational for one to believe, when all one's goals are taken into account, can in principle be at odds with what it is epistemically rational for one to believe, in practice this tends not to happen. There are pressures that tend to keep the two together, which in turn helps explain why in general we feel no need to assess beliefs in terms of how well they promote the total constellation of goals.

Keeping this conclusion in mind, reconsider the complaint that the notion of epistemically rational belief is of little relevance for the everyday assessments we make of each other's beliefs. This much is correct about the complaint: epistemic rationality is concerned with a very specific goal, that of now having accurate and comprehensive beliefs, whereas our everyday assessments are sensitive to the fact that all of us have non-intellectual as well as intellectual goals and long-term as well as short-term goals. Nevertheless, this is not so much a criticism as an acknowledgment that epistemic rationality is a narrowly circumscribed notion. Moreover, its narrowness has advantages. Because the notion is narrowly circumscribed, it can be explicated without reference to other notions of rationality or any of its cognates, and this in turn makes the notion suitable as a theoretical anchor for other notions of rationality, including ones that are less narrowly circumscribed and, hence, potentially more relevant to our everyday intellectual concerns.

The complication, as I have been pointing out, is that the most straightforward way of introducing a derivative notion of rational belief is too crude to be of much relevance for our everyday intellectual concerns. According to the general schema, believing P is rational, all things considered, if it is epistemically rational for him or her to believe that the overall effects of believing P are sufficiently beneficial. However, it is rare for epistemically rational belief and rational belief, all things considered, to come apart in a simple Pascalian manner. There are pressures that keep the two from being in conflict with one another in all but unusual circumstances. So, if non-epistemic goals, values, and needs are to be used in fashioning an account of belief assessment relevant to our everyday practices, they will have to be introduced in a more subtle, indirect way.

A first step is to recognize that our everyday evaluations of each other's beliefs are reason-saturated. We are interested in whether others have been *reasonably* conscientious, *reasonably* careful, and *reasonably* cautious in forming their beliefs. We are also interested in negative assessments, for example, in whether they have been *unreasonably* careless in gathering evidence or *unreasonably* hasty in drawing conclusions from their evidence. The standards of reasonability and unreasonability at work in these assessments are realistic ones. They reflect the fact that all of us have non-intellectual interests and needs, which impose significant constraints on how much time and effort ought to be devoted to intellectual inquiry and deliberation.

Only a realistic notion, which is sensitive to questions of resource allocation, will be capable of capturing the spirit of these everyday evaluations. Indeed, because we

evaluate each other's beliefs in a variety of contexts for a variety of purposes, perhaps several notions will be needed. Still, at least many of our evaluations can be understood in terms of a pair of notions, which I will call "responsible belief" and "non-negligent belief."

As I use the expressions, responsibly believing a proposition is a matter of its being rational to have acquired and subsequently retained the belief, while non-negligently believing a proposition is a matter of its not being irrational to have acquired and subsequently retained the belief. Each of these characterizations itself makes reference to rationality (or irrationality), but this is not an insurmountable difficulty, because the notion of epistemically rational belief can be used as a theoretical anchor to explicate the relevant sense of rationality (or irrationality). The result will be theoretically respectable accounts of responsible and non-negligent belief, that is, accounts that make no ineliminable use of a notion of rationality or any of its cognates.

More specifically, I shall say that one responsibly believes a proposition P if one believes P and one also has an epistemically rational belief[5] that the processes by which one has acquired and sustained the belief P have been acceptable, that is, acceptable given the limitations on one's time and capacities and given all of one's goals, pragmatic as well as intellectual and long term as well as short term. By "processes" I mean to include whatever methods, faculties, and skills are involved in the acquisition and sustenance of the belief. Thus, if an individual has an epistemically rational belief that he or she has spent an acceptable amount of time and energy in gathering and evaluating evidence about P and has also used acceptable methods, faculties, and skills in gathering and processing this evidence, the belief P is a responsible one for the individual to have. Likewise, with respect to a belief Q that has been acquired automatically, without any deliberation, if the individual has an epistemically rational belief that the faculties and skills that generated the belief are acceptable, that is, acceptable in light of all one's goals, the belief Q is responsible.[6]

However, we often do not have a very good idea of how it is that we came to believe what we do. We may not remember or perhaps never knew. Consequently, with respect to many of our beliefs, we may not think that the processes that led to them were acceptable, but by the same token we may not think, and need not have evidence for thinking, that the processes were unacceptable either. Under such conditions, I shall say that the beliefs in question are non-negligent.

More exactly, one non-negligently believes a proposition P if (a) one believes P, and (b) one does not believe, and it is not epistemically rational for one to believe, that the processes by which one has acquired and sustained the belief P have been unacceptable, that is, unacceptable given the limitations on one's time and capacities and given all of one's goals. By "processes" I once again mean to include whatever methods, faculties, and skills are involved in the acquisition and sustenance of the belief. Thus, if an individual does not believe, and it is not epistemically rational for the individual to believe that he or she has spent an unacceptably small amount of time in gathering and evaluating evidence about P, or employed unacceptable methods, faculties, and skills in believing P, then his or her belief P is non-negligent.

The notions of responsible and non-negligent beliefs are less idealized than the notion of epistemically rational belief, and they thus provide a way of acknowledg-

ing that given the relative unimportance of some topics, the scarcity of time, and the pressing nature of many of our non-intellectual ends, it would be inappropriate to spend a significant amount of time gathering information about these topics and deliberating about them. We acquire most of our beliefs with little or no thought at all. I believe that there is a table in front of me because I see it. I don't deliberate about whether to trust my senses. Moreover, in general this is an acceptable way to acquire beliefs. Unless there are concrete reasons for suspicion, it is usually foolish to spend time and effort deliberating about what we are inclined to believe spontaneously. It is better to keep ourselves on a kind of automatic pilot and to make adjustments only when a problem manifests itself.[7] The proverb, don't try to fix what isn't broken, is good intellectual advice as well, and the above notions of responsible and non-negligent belief provide a way of recognizing the aptness of this rule of thumb.

Because most cognitive processing is conducted in a largely automatic fashion, it is important to cultivate intellectual virtues, which serve us well across a range of situations without the need for case-by-case deliberation. An intellectual virtue is a cognitive habit, trait, or skill conducive to producing valuable states of affairs. An epistemic virtue is a special kind of intellectual virtue, one that is conducive to producing epistemically valuable states of affairs. Judiciousness, conscientiousness, and rigor are traits that tend to promote one's having accurate and comprehensive beliefs, and as such they are epistemic virtues. Other epistemic virtues operate more indirectly, in that they are conducive to epistemically valuable states of affairs only when a sufficiently rich array of other epistemic virtues are present. It is arguable, for example, that imagination, creativity, and independence are virtues of this variety.

The cultivation of intellectual virtues is important, I have been pointing out, because most cognitive processing is conducted in a largely automatic fashion. Of course, even if most processes of belief acquisition and revision have to be automatic, we also have good reasons to monitor these processes, keeping an eye out for incongruities, inconsistencies, and the like. Moreover, some issues are so important or so complex that beliefs about them ought not to be formed in a spontaneous or semi-automatic way. They require conscious scrutiny, reflection, and a careful weighing of evidence.

However, only a small fraction of our intellectual methods, practices, and faculties, and only a small fraction of the opinions they generate, can be subject to scrutiny. It is thus necessary to distinguish that which is deserving of attention from that which is not. Making these distinctions is one of the most indispensable and difficult components of leading a well-spent intellectual life. We are constantly confronted with potential intellectual problems. Every new situation presents us with new challenges, if only because we will want to know the best way to react to the situation. We are thus bombarded with potential intellectual projects and questions, but given the total constellation of our goals, some of these projects are more important than others, and likewise, given the scarcity of time, some are more pressing than others. These are the ones most deserving of our attention and time.

The scarcity of time and the urgency of many of our non-intellectual needs have a direct bearing on what we can responsibly believe. One implication, as I have al-

ready mentioned, is that we can have responsible beliefs about unimportant topics even if we have spent little or no time gathering evidence and deliberating about them. Indeed, we can have responsible beliefs about them even if we are in the possession of information which, had we reflected on it, would have convinced us that what we believe is incorrect. This is one of the ways in which responsible belief and epistemically rational belief can come apart. Even if I have evidence that makes it epistemically irrational to believe P, I might nonetheless responsibly believe P, since given the unimportance of the topic and the demands of my other ends, it might not have been appropriate to have taken the time and effort to sift through this evidence.

Similarly, consider the logical implications of a proposition P that is epistemically rational for me. Perhaps not every implication of P will itself be epistemically rational for me, since some may be too complicated for me even to understand. Still, on most accounts of epistemic rationality, a large number of these implications will be epistemically rational for me. But of course, it ordinarily won't be reasonable for me to try to identify and, hence, believe all these propositions. After all, most will be unimportant ones. So, this is another of the ways in which responsible belief and epistemically rational belief can come apart. I can responsibly not believe these propositions even though they are epistemically rational for me to believe.

The standards for responsible beliefs become easier to meet as the significance of the topic decreases, but there are limits. Even with respect to the most insignificant propositions (for example, that the one-pound bag of salt in my pantry contains an even number of grains of salt), I am not free to believe, disbelieve, or withhold judgment as whim dictates. At first glance, it might seem otherwise. If nothing of consequence hangs on which attitude I adopt, it might seem as if any attitude will equally well meet the requirements of responsible belief.[8]

But in fact, this is not so, in part because even if nothing of immediate consequence is at issue, it is difficult to be confident that nothing of long-term consequence hangs in the balance, especially given that beliefs come and go in clusters. Beliefs about trivial matters are not completely isolated from one's other beliefs. They are intertwined with other beliefs and, hence, can have ramifications for one's other beliefs and, in turn, one's decisions about matters which are anything but insignificant. W. K. Clifford makes the point this way:

> If a belief is not realized immediately in open deeds, it is stored for guidance of the future. It goes to make a part of that aggregate of beliefs which is the link between sensation and action at every moment of our lives, and which is so organized and compacted together that no part of it can be isolated from the rest, but every new addition modifies the structure of the whole. No real belief, however trifling and fragmentary it may seem, is ever truly insignificant; it prepares us for more of its like, confirms those which resembled it before, and weakens others; and so gradually it lays a stealthy train in our inmost thoughts which may some day explode into action and leave its stamp on our character forever.[9]

Clifford, in his somewhat breathless style, overstates his case, but he is making an important observation, and it is one that has a bearing on responsible believing. Be-

cause one can never be absolutely sure what information will turn out to be useful and because beliefs about one topic have rippling effects on one's beliefs about other topics, it is possible to have irresponsible beliefs even about matters that seem utterly inconsequential.

Having said this, however, it bears repeating that the standards of responsible belief vary with the importance of the issue under consideration whereas the standards of epistemically rational belief do not. The latter notion is explicated in terms of the epistemic goal, that of now having accurate and comprehensive beliefs, and this goal makes no distinction between important and unimportant truths. There are countless issues about which one might have accurate and comprehensive beliefs, and the notion of epistemic rationality is ruthlessly neutral with respect to them. By contrast, the notion of responsible belief is explicated in terms of the total collection of one's goals, and these goals can be used to distinguish issues of greater and lesser significance. The set of potential issues about which one might have accurate and comprehensive beliefs is thus carved down into a more manageable set of issues about which it is important to have accurate and comprehensive beliefs. These are the issues worthy of one's attention and energies.

The lesson we are left with, obvious as soon as it is stated but not sufficiently acknowledged in the philosophical literature, is that the more important the issue, the more time and effort it is reasonable to devote to having accurate and comprehensive beliefs about it. Accordingly, the standards of responsible believing are more stringent, although they never reach the point where one is required to sacrifice everything else of value in order to have accurate and comprehensive beliefs about the issue in question. Correspondingly, the less important an issue is, the less time and effort it makes sense to devote to it. Accordingly, the standards of responsible believing are less stringent, although they never reach the point at which just anything goes.

The result is a call for moderation. Responsible believing does not require us to be fanatical in our intellectual pursuits. The unreflective life may not be worth living, but neither is the overly reflective life. Time is a scarce commodity, and many of our most important goals are not intellectual ones. However, responsible believing does not permit one to be lackadaisical either. We have good reasons, intellectual and otherwise, to be active in trying to ensure that our belief systems are both accurate and comprehensive. Being a responsible believer requires that one cultivate intellectual virtues, which will serve one well without the need for case-by-case deliberation; it requires that one monitor one's largely automatic processes of belief acquisition and revision; and it requires that one identify and pursue with appropriate diligence those intellectual projects and issues that are worthy of one's explicit attention, although even here it normally does not require one to exercise extraordinary care.

More precisely, responsible believing does not require extraordinary care unless the issue itself is extraordinarily important. The kind of cases I have been emphasizing up until now are ones in which the standards of responsible belief tend to be lower than those of epistemically rational belief, because nothing weighty is at stake. However, when issues of great consequence are at stake, the standards of responsi-

ble belief become correspondingly high. Indeed, they can be more stringent than those of epistemically rational belief. The more important the issue, the more important it is to reduce the risk of error. If, for example, having inaccurate opinions about a given topic would put people's lives at risk, one should conduct especially thorough investigations before settling on an opinion. If one fails to do so, the resulting beliefs will be irresponsible even if they are epistemically rational.

This is possible because epistemically rational belief does not require certainty, not even moral certainty, whereas moral certainty sometimes is required for responsible belief. For a belief to be epistemically rational, one needs to reduce the risk of error to an acceptable theoretical level, that is, to an acceptable level insofar as one's goal is to have accurate and comprehensive beliefs. But the risk of error can be acceptable in this theoretical sense even if one's procedures have been unacceptably sloppy, given that people's lives are at risk. If so, the beliefs in question will be epistemically rational but irresponsible.

The standards of responsible believing vary with the importance of the issue under consideration, and they also vary with one's social role. If it is your job but not mine to keep safety equipment in good working order, the intellectual demands on you to have accurate beliefs about the equipment are more serious than those on me. My belief that the equipment is in good working order might be responsible even if I have done little, if any, investigation of the matter. I need not have tested the equipment, for example. A cursory look might suffice for me, but this won't do for you. The standards of responsible belief are higher for you. You need to do more, and know more, than I.

Of course, even if you have met these higher standards, there are no guarantees that your beliefs will be true. Because it is extremely important for you to have accurate beliefs about the equipment, your beliefs about the equipment won't be responsible unless it is epistemically rational for you to believe that your procedures have been so extraordinarily thorough that they reduce the risk of error to a negligible level. Nevertheless, even extraordinarily thorough procedures can fail, and if so, your beliefs will be responsible even if they are false.

One's social role can also be relevant when the issue at hand is primarily of theoretical interest. For example, my responsibly believing that the principle of conservation of energy is not violated in the beta decay of atomic nuclei is a very different matter from a theoretical physicist responsibly believing this. My familiarity with the issue derives mainly from brief, popular discussions of it. This kind of appeal to authority is presumably enough to make my belief a responsible one, since no more can be reasonably expected of me. More is reasonably expected of the authorities themselves. They are part of a community of inquirers with special knowledge and special responsibilities, and as a result they should be able to explain away the apparent violations in a detailed way.

Social roles play a role in setting the standards of responsible believing because some goals are socially defined. Part of what is involved in being an expert in a field is committing oneself to having well-informed opinions about the central issues of the field, whereas non-experts need not have this as one of their goals and hence need not go to the lengths expected of experts to have informed opinions about these

issues. Because the notion of responsible believing is defined in terms of the total constellation of one's goals, including those that are socially defined, it is able to account for this difference in the standards of responsible believing between experts and non-experts.

In an analogous way, the notion of responsible belief is able to account for the ways in which the standards of responsible believing are shaped not just by one's immediate concerns but also by the kind of life to which one aspires. Because the notion of responsible believing is explicated in terms of all one's goals, including lifetime goals, it builds into the standards of responsible believing the importance of looking beyond momentary concerns and assessing the direction and degree of one's intellectual pursuits (which issues to pursue and how much effort to devote to their pursuit) in the context of what would constitute a worthwhile life. Significant intellectual pursuits require significant time and energy, and hence it is important to weigh the overall benefits of these pursuits against their costs, short term as well as long term, and intellectual as well as non-intellectual. The relative importance of these benefits and costs varies in accordance with the kind of life one wants. Every full, rich human life will involve intellectual projects, if only because one needs to determine effective means to one's non-intellectual ends, but not every full, rich human life will contain the same degree of commitment to intellectual pursuits, which is just to state the banal truth that there are worthwhile human lives that are not the lives of intellectuals. What is not so banal, however, is that the standards of responsible believing for those who have dedicated themselves to an intellectual life are different, and generally higher, than the standards for those who have dedicated themselves to some other kind of life. Or at least this is so with respect to the range of issues (which may be quite broad) to which the intellectual has dedicated himself or herself. The moral, which is an instance of the point made earlier about the standards of responsible belief in a field of inquiry being more demanding for experts than non-experts, is that an intellectual life brings with it special responsibilities, and just as one would expect, the responsibilities are principally intellectual in nature.

The most general and important point to emphasize here, however, is that responsible believing requires that one's intellectual pursuits be appropriately situated in the context of the kind of worthwhile life one wants for oneself. Situating intellectual pursuits in the context of one's overall life, in turn, requires one to take seriously that one's life is going to extend into the future. Part of what is involved in being a responsible believer is having a sense that one's life is temporally extended, even if one cannot be certain how far into the future it will extend. One's current activities are to be assessed in the context of what one wants one's life as a whole to be, or at least in terms of what one wants the remaining part of one's life to be. Seeing the current moment as one moment of a temporally extended life is as important for one's intellectual pursuits as it is for one's other pursuits, and the notion of responsible belief has the importance of these considerations built into it. The standards of responsible believing vary not only in accordance with the importance of the issue under consideration and in accordance with the social role of the inquirer, it also varies in accordance with the kind of life to which the inquirer is committed.

The above notion of responsible belief has other benefits built into it as well. For

example, it provides the underpinnings for our everyday evaluations of decisions and actions. These evaluations, like our everyday evaluations of beliefs, are sensitive to issues of resource allocation. They reflect the fact that we do not have time to deliberate over each and every decision. Here too we have no choice but to operate in a largely automatic fashion. Typically, we deliberate only when there are special reasons to do so; for example, when the issues are especially important or when our accustomed ways of acting begin to create unexpected problems.

However sensible this practice may be, it sometimes results in our doing things that we ourselves would have regarded as unwise, had we taken more time to think about them. Even so, we need not be acting irresponsibly. After all, given the constraints on our time and given the importance of our other needs and goals, there may not have been any reason for us to suspect that special care was needed, and, accordingly, we may have responsibly believed that our actions would not lead to unacceptable results. The notion of responsible belief thus provides part of the necessary infrastructure for an account of responsible action. At least in the most simple cases, an individual has acted responsibly if he or she responsibly believed that his or her action would yield acceptable results.

The importance of the notions of responsible and non-negligent belief for our everyday purposes does not detract from the importance of the more narrowly circumscribed and idealized notion of epistemically rational belief. On the contrary, the former put us in a better position to understand the significance of the latter. Our everyday evaluations of beliefs and actions are reason-saturated; they presuppose notions of rationality, reasonableness, or one of their cognates. They thus leave us within the circle of terms for which we want a philosophical account. The notion of epistemically rational belief permits an escape from this circle. It is thus indispensable for an adequate philosophical understanding of these assessments; it is indispensable for a complete theory of rationality.

An appreciation of this point is the key to responding to the common complaint that the issues and concepts of contemporary epistemology are too rarified and too far removed from our actual intellectual practice and concerns to matter. Descartes and many of the other great epistemologists of the modern period regarded epistemology as the czar of the sciences. Its role was to provide assurances of the reliability of properly conducted inquiry, and thus to place science, and inquiry in general, on a secure foundation. This conception of epistemology is almost universally rejected by contemporary epistemologists. But then, why is epistemology important? Is its role nothing more significant than that of clarifying our concepts of knowledge and rational belief? My answer is that epistemology is important because it has a foundational role to play, but not that of a guarantor of knowledge. Its role, rather, is that of providing a philosophically respectable foundation that can be used to understand the everyday notions we employ to assess each other's beliefs and actions. Its role, in other words, is the less flamboyant but nonetheless crucial one of providing a theoretical anchor for a general theory of rationality.

Notes

1. See Richard Foley, *Working Without Net* (New York: Oxford University Press, 1993), especially pp. 3–15.

2. The expression "estimated desirability" is Richard Jeffrey's; see *The Logic of Decision*, 2nd ed. (Chicago: University of Chicago Press, 1983).

3. Recall that the basic schema governing the rationality of plans, decisions, actions, etc., is concerned with rationality in the abstract. A plan can be rational in this sense even if it is not chosen or chosen for the wrong reasons. The same point applies to the above schema for epistemically rational belief. To use the terminology introduced by Roderick Firth, the above schema is concerned with epistemically rational belief in the propositional sense as opposed to the doxastic sense. It can be epistemically rational for one to believe *P* in the propositional sense even if one does not believe *P* or believes *P* for poor reasons. In the case of the latter, the belief *P* is propositionally rational (since one has good reasons to believe *P*) but not doxastically rational (since it was not those good reasons that led to the belief). See Roderick Firth, "Are Epistemic Concepts Reducible to Ethical Ones?" in *Values and Morals*, A. Goldman and J. Kim, eds. (Dordrecht: Reidel, 1978), 215–29. For a reliabilist treatment of this distinction, see Alvin Goldman, *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986), 112.

4. According to probabilists, when assessing the rationality of opinions, the primary phenomena to be assessed are not beliefs *simpliciter*, but rather degrees of beliefs (or subjective probabilities). Nevertheless, the above point holds for probabilists as well, *mutatis mutandis*. Only when one has degree of belief *x* in proposition *P* and this degree of belief meets the recommended probabilist conditions is it epistemically rational for one to have *x* degree of belief in the proposition that this opinion contributes to the goal of having accurate and comprehensive degrees of belief. Accuracy here is to be construed in terms of one's subjective degrees of belief conforming to the objective probabilities. See Bas van Fraassen, "Calibration: A Frequency Justification for Personal Probability," in *Physics, Philosophy, and Psychoanalysis,* R. S. Cohen and L. Laudan, eds. (Dordrecht: Reidel, 1983), 295–319. See also Richard Foley, *Working Without a Net*, especially pp. 156–57.

5. Recall that 'epistemically rational belief' refers to a belief that is propositionally rational. Thus, the belief in question here, namely, the belief that one has acquired belief *P* in an acceptable way, need not be doxastically rational (see note 3).

6. Note that there is a distinction between one's being responsible for one's behavior, in the sense of having some degree of control over it, and one's behaving responsibly, in the sense of one's having behaved in a balanced, choiceworthy way. The contrasting notion to the latter is behaving irresponsibly, where the contrasting term to the former is behaving in a non-agency kind of way, for example, compulsively. A similar distinction applies to responsible belief, and the above account is intended as an account of responsible belief in the latter sense, such that the contrasting notion is irresponsible belief. Of course, many would argue that the former sense of responsible belief (behavior) is a prerequisite of the latter sense; that is, a belief (behavior) is neither responsible nor irresponsible in the latter sense unless one is responsible for it in the former sense. If so, then the above account should be read as presupposing an appropriate degree of agency over one's beliefs.

7. See Kent Bach, "Default Reasoning," *Pacific Philosophical Quarterly* 65 (1984): 37–58.

8. I owe Abrol Fairweather for this point.

9. Clifford, W. K., "The Ethics of Belief," in *Theory of Knowledge*. Louis Pojman, ed. (Belmont, Calif.: Wadsworth, 1993), 502–505.

14

EPISTEMIC OBLIGATION AND THE POSSIBILITY OF INTERNALISM

Hilary Kornblith

S he shouldn't have believed him." "You shouldn't believe everything you read." "He should have known better." This way of talking about belief is altogether common. When people say this sort of thing, it seems, they quite often say things that are true. But this talk of epistemic obligation can seem quite puzzling, and it is puzzling in a way that talk of obligations to act is not. If I can't be heard in the back of the room when I'm presenting a paper at a conference, then I may be criticized for not having spoken louder: I should have spoken louder, it will be said, and rightly so. I am criticizable here, it seems, because I could have spoken louder but didn't; how loud I speak is subject to my voluntary control. But as William Alston[1] and Alvin Plantinga[2] have pointed out, believing surely seems different here. I don't have voluntary control over my beliefs. Although I can simply decide to speak louder, I can't simply decide to believe or disbelieve. Alston and Plantinga draw the obvious conclusion: talk of epistemic obligation—what they call the deontological conception of justification—is thereby undermined.

Now those who accept the view that belief is subject to direct voluntary control will be entirely unperturbed by this argument, for epistemic oughts are threatened only on the assumption that we reject a voluntarist conception of belief. And there are some who wish to endorse such a voluntaristic view. I myself, however, like many others, do not believe that voluntarism about belief is true, and for the purposes of this essay, I will simply suppose that we do not have direct voluntary control over our beliefs. Are Alston and Plantinga right in thinking that talk of epistemic oughts will then need to be rejected as well?

In several recent essays, Richard Feldman[3] has attempted to defend the legitimacy of talk of epistemic obligation even on the assumption that doxastic voluntarism is false. I want to examine Feldman's defense of epistemic deontology in some detail. I believe that Feldman is right in thinking that a commitment to talk of epistemic oughts is fully compatible with the rejection of doxastic voluntarism. But Feldman wishes to use his defense of the legitimacy of epistemic deontology as a springboard for a particular account of our epistemic obligations. Feldman is an ev-

identialist: he holds that we ought to believe only those propositions that our evidence supports. But Feldman's own defense of the legitimacy of epistemic obligations does not leave room for his defense of evidentialism, or so I shall argue. In the end, Feldman's views about epistemic obligation suggest an altogether different defense of internalism than the one which he himself wishes to endorse. I will argue that the most defensible version of internalism is committed to an interesting and controversial empirical research program.

I

Feldman once suggested that we may understand epistemic obligation on the model of contractual obligation. When Mary signs a contract, she thereby undertakes certain obligations; should she find herself unable to fulfill those obligations, this does not show that she was therefore under no obligation to comply with the contract. Thus, to take one of Feldman's examples, when I bought a house, I signed a contract with the bank; I am obligated to pay a certain amount of money each month. If I find myself unable to pay, that doesn't show that I am not obligated to pay nonetheless. Similarly, one may legitimately talk of an individual having epistemic obligations even if that individual is unable to comply with them. Contrary to Alston and Plantinga, Feldman argued, it is perfectly legitimate to talk of obligations even in cases in which the ability to comply with them is absent. Thus, the fact that belief is not subject to our voluntary control does not undermine the suggestion that we are obligated to believe in certain ways.

But Feldman now rejects the analogy between epistemic and contractual obligation, and, to my mind, for good reason. As Feldman now comments, "It's difficult to see any basis for saying that we all have a contractual obligation to believe things. Surely no such contract is explicit . . ." As Feldman also points out, there seems no ground for suggesting that we have implicitly taken on any sort of contract to believe things either.[4]

Feldman now sees epistemic obligations as a kind of "role ought."

> There are oughts that result from one's playing a certain role or having a certain position. Teachers ought to explain things clearly. Parents ought to take care of their kids. Cyclists ought to move in various ways. Incompetent teachers, incapable parents, and untrained cyclists may be unable to do what they ought to do. Similarly, I'd say, forming beliefs is something people do. That is, we form beliefs in response to our experiences in the world. Anyone engaged in this activity ought to do it right. In my view, what they ought to do is to follow their evidence (rather than their wishes or fears). I suggest that epistemic oughts are of this sort—they describe the right way to play a certain role.[5]

By seeing epistemic obligations as role oughts, we see that Alston and Plantinga are mistaken in thinking that the legitimacy of talk of epistemic oughts presupposes some kind of doxastic voluntarism. Thus, Feldman claims, "Even in cases in which

a believer has no control at all, it makes sense to speak of what he ought to believe and ought not to believe."[6]

Now Feldman has some insightful things to say about the relationship between playing a certain role and the role obligations to which it leads. Role "oughts are not based on what's normal or expected. They are based on what's good performance."[7] But what is good performance?

> What counts as good performance in a role, and thus determines how a role ought to be carried out, may be dependent in certain ways on what people are generally able to do. Consider, for example, the claim that teachers ought to explain things clearly. Arguably, what counts as a clear explanation is dependent at least in part on what people are able to say and what people are able to understand. One could imagine standards for clear explanation that are so demanding that no one could ever meet them. It is not true that teachers ought to explain things that clearly. Similarly, the standards of good parenting or good cycling that apply to us are not at super-human levels. It's not true that parents or cyclists ought to do things that would require them to exceed the sorts of capacities people have. It is consistent with this, however, that an individual ought to do things that he or she is not able to do. . . . Thus, even if the standards for good performance in a role are in some way limited by the capacities of those who fill the role, it is not the case that the existence of those standards implies that individuals must have basic or non-basic voluntary control over that behavior that is judged by those standards.[8]

What Feldman says here, it seems to me, is exactly right. On the one hand, he denies that talk of epistemic obligations requires voluntary control over our beliefs, contrary to Alston and Plantinga; on the other hand, since role obligations are tied to ideals of good performance, and what counts as good performance for a human being is in part determined by human abilities, the content of our epistemic obligations is not wholly independent of what we are capable of either. This sensible middle ground is the natural home of epistemic deontology. What it does not leave room for, however, is the defense of the substantive account of our epistemic obligations that Feldman prefers.

First, consider a strong version of evidentialism, what Feldman calls O1:

> For any proposition p, time t, and person S, S epistemically ought to have at t the attitude toward p that is supported by S's evidence at t.[9]

In order to fill this out, Feldman will need to give us an account of what it is for one's evidence to support a proposition. But the problem here is largely independent of how this account is filled out. For on any account of what it is for one's evidence to support a proposition, so long as the notion of evidence itself is specified in ways that are independent of human abilities, there will be propositions supported by one's evidence where the connection between evidence and proposition supported is extraordinarily elaborate. For example, on most accounts of what it is for evidence to support a proposition, everyone who has ever existed has had sufficient evidence to support belief in the incompleteness of arithmetic. Never mind that it took the ge-

nius of Gödel to figure this out; there is simply an objective relationship between our evidence and the proposition that arithmetic is incomplete. O1 thus entails that all individuals ought to believe in the incompleteness of arithmetic.

This is not, of course, just a problem for apparently a priori claims like the incompleteness of arithmetic. In the case of claims for which we have empirical evidence, so long as one's conception of evidence is objective and specified independent of human abilities, there will be propositions supported by one's evidence that will require the genius of a Gödel, or perhaps a degree of genius which simply is not to be found among human beings, in order to recognize that the connection between evidence and that for which it is evidence actually holds. O1 entails that individuals who have the relevant evidence ought to believe the claims which the evidence supports, however elaborate the connection between evidence and supported claim; indeed, even when the connection is so elaborate that it exceeds human capabilities, O1 entails that anyone who has the relevant evidence is obligated to believe the proposition supported by it.

Such a view is clearly in conflict with the understanding which Feldman now offers of epistemic obligations as role oughts. Indeed, Feldman's insightful comments about the ways in which human limitations thereby limit what a person occupying a role is obligated to do mark a dramatic shift from the position he and Earl Conee defined in their essay, "Evidentialism." There, in considering views of epistemic justification that explicitly take account of human limitations, they comment,

> There is no basis for the [claim] that what is epistemically justified must be restricted to feasible doxastic alternatives. It can be a worthwhile thing to help people to choose among the epistemic alternatives open to them. But suppose that there were occasions when forming the attitude that best fits a person's evidence was beyond normal cognitive limits. This would still be the attitude *justified* by the person's evidence. If the person had normal abilities, then he would be in the unfortunate position of being unable to do what is justified according to the standard for justification asserted by [evidentialism]. This is not a flaw in the account of justification. Some standards are met only by going beyond normal human limits. (19)

Clearly enough, the account Feldman now provides of epistemic obligations as role oughts is incompatible with the evidentialist account of our epistemic obligations that he once offered.

Feldman's current account of our epistemic obligations is slightly weaker than O1; instead, he currently endorses the following view (O2):

> For any person S, time t, and proposition p, if S has any doxastic attitude at all toward p at t and S's evidence at t supports p, then S epistemically ought to have the attitude toward p supported by S's evidence at t.[10]

This is weaker than O1 since it does not entail that individuals are obligated to believe anything at all; it only entails that where one does have beliefs, one is obligated to believe in accord with one's evidence.

But this does not seem to solve the problem indicated above. The account of one's epistemic obligations defined here is entirely independent of human abilities; the account of role oughts specifies our obligations as believers by way of our cognitive limitations. The upshot seems clear: One may either be an evidentialist, and specify epistemic obligations in ways independent of human capacities, or one may explain epistemic obligations in terms of role oughts, and then explicitly take account of human capacities; but one may not consistently combine both of these views.

Although Feldman does not provide us with an account of evidential support, and evidentialism is fundamentally incomplete without it, Feldman now addresses the kind of problem just raised in a way that speaks to the fundamental shift in his position since its initial statement in the essay with Conee. Feldman comments,

> The evidentialist account of what we ought to believe relies crucially on the notion of evidential support. Analyzing this notion in a fully satisfactory way is no easy task. Among the problems to be worked out is that of determining which logical consequences of a body of evidence are supported by that evidence. There are possible cases in which a person has evidence that implies some proposition, but the connection between that evidence and that consequence is distant and difficult to see. It may be well beyond the intellectual talents of the person. I believe that in such cases the person ought not to believe the consequence. Given his failure to see that it is a consequence, to believe it (barring other reasons to believe it) would be rash. Furthermore, as I understand (O2), it has exactly the right result in this sort of case. The fact that a person's evidence implies some proposition is not sufficient for the evidence to provide evidential support for the proposition. Roughly, only those propositions whose connection to the evidence the person apprehends are actually supported by his evidence. And I think ascertaining this connection is itself an element of the person's evidence.[11]

But this suggestion, I believe, simply muddies the waters. First, if Feldman really means to suggest that however strong an individual's evidence for a proposition, that evidence does not provide "evidential support" unless the individual also apprehends the connection between the proposition and its evidence, then evidentialism loses its power to criticize individuals for failing to recognize the evidential relations among their beliefs even in cases where the evidential connection is well within the cognitive powers of the individual.[12] One would have thought that the entire point of evidentialism was precisely that there are certain evidential relations among our beliefs whether we recognize them or not, and we ought to recognize them. Feldman not only spoke this way in motivating evidentialism in his early essay with Conee, he speaks this way now as well. Thus, in "Voluntary Belief and Epistemic Evaluation," he says,

> In my view, the right way to carry out one's role as a believer is to form beliefs that are supported by one's current evidence. That is, if one is considering a proposition, then one ought to believe it if one's evidence supports it, ought to disbelieve it if one's evidence goes against it, and ought to suspend judgment if one's evidence is neutral. (20–21)

The same view is endorsed in his essay, "The Ethics of Belief" (9). There is a distinction that is relevant here which Feldman and Conee appealed to in their early essay, between having evidence and using it,[13] and this distinction, it seems to me, is one which evidentialists should want to hold on to. One may have evidence that one fails to use, and indeed, that one fails to apprehend, but one should not therefore insist that one's evidence fails to provide evidential support simply because one fails to recognize that it bears the appropriate relation to the proposition it is evidence for. Feldman's more recent discussion of this issue seems to run together issues which he and Conee rightly kept apart.

Indeed, consider the agent who has evidence E which entails the truth of a certain proposition p and, on considering the relationship between E and p, forms the belief that E supports p. What will the evidentialist say about this proposition, that E supports p? One may certainly come to believe propositions like this without adequate evidence, and so an evidentialist ought to want to know whether this proposition itself is supported by adequate evidence. On the view advanced in his essay with Conee, this question is easily resolved. In this case, the claim that E supports p is itself supported by the agent's evidence, since E entails p. But on the new view being advanced now, which claims that one has evidential support only if there is some apprehension that the evidence supports p, we seem headed for an infinite regress: each layer of evidential support can do its required work only if there is apprehension that it does so, which, in turn, must itself be justified by appropriate evidential support, requiring still further apprehension, ad infinitum.[14] The root idea of evidentialism, so nicely articulated in the essay with Conee, avoided this problem altogether.

Finally, it is worth pointing out that, contrary to Feldman's suggestion, the move from (O1) to (O2) does nothing at all to suggest that evidence only provides evidential support if it is believed to provide such support. Remember what (O2) states:

> For any person S, time t, and proposition p, if S has any doxastic attitude at all toward p at t and S's evidence at t supports p, then S epistemically ought to have the attitude toward p supported by S's evidence.

If there is reason to read the notion of "epistemic support" in the way Feldman now suggests, then it should, presumably, be read that way in (O1) as well. The question of how to read the notion of "epistemic support" is entirely independent of the question of whether we should move from (O1) to (O2). Nothing in (O1) precludes the new reading of "epistemic support," and nothing in (O2) encourages it.

I conclude that the initial defense of evidentialism presented by Conee and Feldman is inconsistent with the suggestion Feldman now makes that epistemic oughts are role oughts. Feldman's attempt to modify the initial account of evidentialism by requiring that a body of evidence only supports a proposition for an individual if that individual apprehends that it does is unmotivated and incoherent. We are thus left with a choice between the original formulation of evidentialism, which characterized evidence in a way which is entirely independent of human ability to recog-

nize evidential connections, and the characterization of epistemic oughts as role oughts, which specifies our duties in ways that crucially depend on human limitations and capacities.

II

Is Feldman right in thinking that epistemic obligations are role oughts? I don't think that he is, although I believe that much of what Feldman says about role oughts is in fact true about epistemic obligation, or at least, about one kind of epistemic obligation.

When Feldman discusses examples of role obligations other than epistemic obligations, he mentions the roles of being a teacher, being a parent, and being a cyclist. But there are lots of different kinds of roles that individuals might occupy, and if we are properly to understand what the force of role oughts comes to, we will need to look at other roles one might occupy as well. One might, for example, occupy the role of being a tyrant, a con artist, or a thief. There are standards of good performance that go along with each of these roles, and one might therefore speak of certain role oughts that flow from each of them. Thus, if one wants to be a good tyrant, perhaps one ought to be particularly brutal. Nevertheless, we would not say, even of someone who did in fact want to be a good tyrant, that he ought to be particularly brutal. Rather, it seems, he ought to stop wanting to be a tyrant. This contrasts with the epistemic case, where we not only want to say that if someone wants to be a good believer, he or she should believe in certain ways; we also wish to endorse the claim that individuals ought, without qualification, to believe in those ways which, as a matter of fact, flow from good performance of the role of being a believer. Since being a tyrant or a con artist or a thief is just as much of a role as being a believer, what is the relevant difference here that allows us to detach the ought judgment in the case of believers, but prevents us from detaching the ought judgment in the case of tyrants, con artists, and thieves?

Now one point that Feldman makes about being a believer is that it is not in any way optional:

> it is plausible to say that the role of a believer is not one that we have any real choice about taking on. It differs in this way from the other roles mentioned. It is our plight to be believers. We ought to do it right.[15]

Now Feldman is surely right that we have no choice about being believers, and in this respect it is different from typical cases of tyrants, con artists, and thieves. And Feldman is also right, I believe, in saying that, when it comes to one's role as believer, one ought to do it right. But the fact that we have no choice about being believers has nothing to do with why it is that we should do it right. Many people are forced into horrible roles; they are put in positions over which they have no choice. Some are forced into slavery; others into prostitution. Much as they may have no choice about

playing certain roles, we don't want to say in these cases that, whatever the role, they ought to perform them well. While it is true that we ought to perform our roles as believers well, it is not simply because we cannot help but play such a role.

Epistemic oughts thus seem importantly different from role oughts. If role oughts simply "describe the right way to play a certain role," then, qua role ought, they carry virtually no normative force. Moreover, in cases where someone has no choice about playing a role, this imparts no greater normative force than the role ought would have had were the role optional. But epistemic oughts do carry a good deal of normative force. As Feldman rightly says, when it comes to believing, we ought to do it right. An account of the nature of epistemic oughts should explain why this is so. The suggestion that they are role oughts does not do so.

Although I believe that Feldman is therefore mistaken in thinking that epistemic oughts are role oughts, much of what he says about role oughts offers important illumination on the subject of epistemic oughts. Thus, when Feldman argues that standards of good performance in a role must in some ways take account of human capacities, and yet, at the same time, "it is not the case that the existence of those standards implies that individuals must have basic or non-basic voluntary control of that behavior that is judged by those standards,"[16] I believe that what Feldman says applies not only to standards of good performance in roles, but to ideals in general. An appropriate human ideal must in some ways be responsive to human capacities. Ideals are meant to play some role in guiding action, and an ideal that took no account of human limitations would thereby lose its capacity to play a constructive action-guiding role. At the same time, our ideals cannot be so closely tied to what particular individuals are capable of that we fail to recognize that some individuals at some times are incapable of performing in ideal ways. There is a large middle ground here, and it is here that reasonable ideals are to be found.

Moreover, once we recognize that our ideals must lie somewhere within this large middle ground, we see that the defensibility of the oughts that flow from our epistemic ideals does not require the level of voluntary control over our beliefs that Alston and Plantinga insist upon, nor does it leave room for epistemic oughts that are as detached from human capacities as Feldman and Conee once urged. Ideals that are wholly detached from human capacities can play no action-guiding role; they provide us with nothing to aspire to. Epistemic ideals of this sort would fail to make sense of the interest of epistemological theorizing. We care so deeply about epistemic ideals precisely because we wish to realize, or at least, approximate them. But ideals wholly divorced from human capacities cannot engage us in this way. Conversely, ideals which are so constrained by individual capacities as to obey narrow versions of the "ought implies can" principle set our standards too low; there must be room, in any reasonable ideal, for the thought that at least some of us, at least some of the time, are not able to do what we ought.[17]

Unlike role oughts, there are not ideals answering to every role which human beings might occupy. Although a teacher might reasonably speak of a conception of what the ideal teacher might do, and thus aspire to behave in those ways, we may also coherently say that certain roles simply do not answer to any legitimate conception of a human ideal. Although the role of being a slave might be performed well,

it is perfectly reasonable to suppose that being a slave, whether the role be performed well or badly, is no part of any acceptable human ideal. It is for this reason that oughts which flow from human ideals have a degree of normative force that is not shared by role oughts.

A defensible ideal, one that may serve as a viable ground for a deontological conception of epistemic justification, thus requires a certain sort of empirical defense. One must show that there is reason to believe, at a minimum, that the ideal occupies what I have been calling the large middle ground between those ideals that are wholly insensitive to human capacities, thereby setting standards that are impossibly high and cannot perform their action-guiding role, and those that are so constrained by what an agent can do that they set standards so low as to provide an ideal unworthy of our pursuit. The empirical defense that our epistemic ideals thus demand will entail at least some real understanding of human cognitive capacities. A deontological conception of epistemic justification will thus need to seek aid from cognitive science.

This is not a conclusion that will come as entirely welcome to many defenders of deontological conceptions of justification. Remember that Feldman and Conee presented their initial conception of evidentialism in a way that insulated it from the need for any such empirical defense. If human capacities are as irrelevant to our epistemic ideals as Feldman and Conee there suggested, then no empirical defense of those ideals is called for. Ironically, however, if I am right in thinking that deontological conceptions of justification do in fact owe us the kind of empirical defense I sketched above, then there may be more hope for Feldman and Conee's original view than I have thus far suggested.

I have been pointing out that the *defense* which Feldman and Conee offered of their original view does not sit well with the remarks Feldman now makes about epistemic oughts and the ways in which they need to be responsive, however indirectly, to human cognitive capacities. This does not show, however, that the version of evidentialism that Feldman and Conee once defended by a priori means might not find a substantive empirical defense when confronted with the facts about our cognitive capacities. Indeed, more than this, evidentialism, as it currently stands, is more a theory sketch than it is a theory; until we know a good deal more about how the notion of evidence is to be filled out, evidentialism is extremely difficult to evaluate. The empirical constraints I am now endorsing on deontological conceptions of justification, however, may do far more than serve to defend the theory sketch that Feldman and Conee have offered; the empirical constraints may serve to fill in some of the details, providing some additional meat on their bare-bones analysis.

Let me explain how this might work. There is a debate within the cognitive science community right now about whether human beings actually reason in ways that employ content-independent formal principles. On one side are the mental logicians, who believe that certain logical principles are innate and that they guide a good deal of our inference.[18] We do, of course, make performance errors. There are factors that may interfere with the smooth working of our inferential mechanisms. But if this view is right, then an important component of human inference is a logic module that guides our reasoning by way of a priori content-neutral principles.[19]

Opposed to the mental logicians are a group of cognitive scientists who see human inference as primarily domain-specific.[20] On this view, the innate inferential mechanisms that underlie our patterns of reasoning are ones that operate only in certain content domains, providing us with reliable inference within those domains. On the mental logic view, the appearance of reasoning that is contrary to the preferred logical principles is explained away as a product of interfering factors. On the view of human reasoning as domain-specific, the appearance of content-neutral logical reasoning is explained away as a byproduct of reliable content-sensitive inferential strategies.

Suppose for a moment that the mental logicians are right: we are innately endowed with a mental module that is hard-wired with a powerful set of principles of inference. Now this does not mean, of course, that we are capable of actually using these inferential principles in such a way as to generate any arbitrary instance of their deductive consequences. In practice, there may be many different sorts of factors that prevent us from actually seeing some of the consequences of a particular belief. Consider an analogy with the theory of grammar. Noam Chomsky argues that competent speakers of English have a complete grammar of the language formally represented in their minds, and this formal grammar guides the linguistic behavior of those speakers. Under the guidance of this grammar, competent speakers rightly recognize that "The man walked to the store" is grammatical and "Walked store man the to" is not. At the same time, there are strings of words that are in fact grammatical according to the rules of the stored grammar, but which will not be recognized in practice by almost any speaker of English. "The man the dog my mother bought bit died" is, mistakenly, thought to be ungrammatical by most competent speakers of English. More than this, there are grammatical strings of words which, in practice, could not be recognized by any human being, if only for the reason that any human being would die before the end of the string was reached. Nevertheless, there are good reasons for adopting an idealization under which we speak of these strings as grammatical. If our interest is in our grammatical competence itself, and not the way in which our power to draw on that competence is limited by other factors, such as the length of a human life, we will be forced to speak of strings of words as grammatical that no human being could recognize as such. While this idealization exceeds the actual capacities of human beings in a number of important respects, and, accordingly, fails to fit with their actual performance, it is nevertheless grounded in an empirically based understanding of human competence.

Now one might adopt a similar attitude toward our reasoning competence, and if the mental logicians are right, and if the logic module is loaded with a sufficiently powerful formal system, our reasoning competence would be quite substantial. There is as much motivation for talk of reasoning competence here as there is in talking of grammatical competence in linguistic theory. Suppose, in particular, that our reasoning module includes formal principles powerful enough to generate a proof of the Gödel incompleteness result. Now most human beings are simply incapable of following, let alone discovering, this particular proof. But it could be that what these mere mortals lack is certain powers of concentration, say, which Gödel, and very few others, enjoy. If this were true, then there would be a natural idealiza-

tion of human reasoning competence that would dictate that the Gödel result is within the scope of human competence even though, in practice, there are very few individuals who are capable of recognizing it as true. And, of course, the same would be true of certain claims that exceed the abilities in practice of all human beings, just as is in the case of grammar. Recognizing this, we might be inclined to say, as Feldman and Conee did, that the standards of epistemic justification, now interpreted as dictated by our reasoning competence, will sometimes exceed the powers — suitably understood — of all human beings. Such a conclusion, as Feldman and Conee once said, does not reflect badly on the standards of epistemic justification; it merely forces us to see that human beings fall short in certain respects. The standards of epistemic justification, standards dictated by our reasoning competence, determine what we ought to believe.

Now I am not suggesting that we ought to reason in accord with whatever principles happen to underlie our actual reasoning.[21] But if the principles underlying our actual reasoning were of the sort that the mental logicians suggest, then there would be good reason for seeing those principles as defining not only our reasoning competence, but the standards which, ideally, our reasoning ought to meet.

A fully spelled-out evidentialism will need more than just a theory of ideal reasoning; it will need an account of the ultimate input to our reasoning processes as well. Here too, we may draw on work in cognitive science to fill out the account. Some evidentialists favor an account in which claims about sense experience serve a foundational role; others allow claims about physical objects to do this work. Empirical work on the nature of input processes can be drawn upon here to resolve this issue.

Notice that if the mental logicians are correct, the resulting account of our reasoning competence — whatever the account of input systems which accompanies it — will look a good deal like the kind of account favored by many internalists about justification. The account will look like a particularly strong version of internalism if the input systems include some kind of introspection that affords access not only to input states, but to features of the reasoning process itself. Whether internalism is true, and if so, what kind of internalist account we should favor, thus becomes an empirical question, and a tractable one at that.

Consider what would be true if the mental logicians are mistaken, and the kind of reasoning that humans embody consists primarily of domain-specific reasoning strategies. In this case, an idealized human reasoning competence will not look anything like the evidential principles favored by evidentialists, for none of them will have the requisite a priori status. The connection between propositions justified and the evidence for them will be de facto reliable, on this view, but the reliability will trade on contingent features of the world and the way in which our inferential processes exploit them. Even assuming a very generous account of what we are capable of introspecting, this account will not bear any similarity to internalist accounts of justification. If an account like this were true, externalism would be vindicated.

There are other possibilities as well. I have been assuming, both in the case of the mental logic view and in the case of the domain-specific reasoning view, that an ide-

alized view of our reasoning competence also looks like a normative view we would want to endorse. This is a very charitable assumption, and things needn't work out quite so neatly. Important elements of our idealized reasoning competence may embody standards that no sane normative theory of reasoning would endorse. If we are to endorse some kind of deontological theory of epistemic obligations, there will need to be some way of getting from our idealized competence to a set of belief-forming processes worthy of enshrining as a normative ideal. Without such a route from our competence to our normative ideal, talk of epistemic duties or obligations surely is undermined. This is just another way of putting Feldman's recent point about the way in which an account of good performance must answer to the limitations on human capacities.

My attempt to defend the potential viability of evidentialism and internalism, and, in particular, my suggestion that the viability of these views ultimately turns on certain empirical questions about human psychology, will probably find few sympathetic ears among evidentialists and internalists. "With friends like this," they may well say, "who needs enemies?" I understand this reaction. After all, one important strand in internalist and evidentialist thought, at least in this century, involves a deep-seated commitment to epistemology as first philosophy, and a view of philosophy as wholly independent of empirical questions. But although the terms 'internalism' and 'evidentialism' are recent arrivals on the philosophical scene, the views themselves are not, and the idea that these views might embody substantial psychological commitments is in no way a new idea, imposed from without. Descartes certainly had substantive views about the powers of introspection, without which his epistemology could not get off the ground; that Locke and Hume and Kant had similarly substantive commitments to claims about human psychology is equally clear. These philosophers did not think that epistemology was somehow independent of psychology. The commitment of these philosophers to internalism was in fact based in their psychological commitments. In this respect, I believe that the internalism and evidentialism of old is an improvement on its most recent de-psychologized incarnations. What I am urging is that the most viable forms of internalism and evidentialism are of this more old-fashioned, psychologically engaged sort.

III

One of the advantages of viewing internalism in this way is that it allows internalists to respond to Alvin Goldman's recent critique, "Internalism Exposed."[22] Goldman sees the heart of internalism as generated by a commitment to what he calls "the guidance-deontological conception of justification," or simply, "the GD conception." The GD conception marries two appealing ideas. First, a historically important motivation for engaging in epistemological theorizing, and, indeed, more than this, a philosophically important motivation for engaging in epistemological theorizing, is the idea that an adequate epistemological theory would guide the concerned epistemic agent in the conduct of inquiry. We are interested in epistemology precisely because we desire to improve our epistemic performance; an ad-

equate epistemology ought to tell us how to achieve such improvement. Second, the deontological conception of justification is just the idea that we have certain epistemological duties, obligations, or responsibilities. Precisely what these are will vary with different ways of filling out the deontological view, but, depending on one's chosen deontology, we are either obliged to follow the evidence, or to seek out appropriate evidence, or to attend to the evidence, or to cultivate appropriate intellectual habits, or to be motivated in our epistemic endeavors in appropriate ways. Each of these ideas is appealing, and, as Goldman points out, they are clearly related. Any way of filling out the deontological conception, it seems, will guide us in our epistemic activities.

Goldman's critique of internalism proceeds by introducing two additional claims:

> [First, a] certain constraint on the determiners of justification is derived from the GD conception, that is, the constraint that all justification determiners must be *accessible to*, or *knowable by*, the epistemic agent. [And second, the] accessibility or knowability constraint is taken to imply that only internal conditions qualify as legitimate determiners of justification. So justification must be a purely internal affair. (272)

Goldman then argues, at length, and, to my mind, persuasively, that the accessibility requirement is untenable.

What are the implications of this, according to Goldman? Note that how direct an implication this has for internalism may depend, in part, on the connection between the GD conception and the accessibility requirement. There can be little doubt that many internalists have argued, precisely as Goldman suggests, from some version of the GD conception to one or another version of the accessibility requirement. And for this reason alone, showing, as Goldman does, that the accessibility requirement is untenable does important philosophical work. But does the GD conception itself actually force one to adopt some kind of accessibility requirement? Interestingly, Goldman says that it does not:[23]

> The argument from the GD conception of justification to internalism does not work. Internalism can be derived only from a suitably qualified version of the [accessibility] constraint because the unqualified version threatens to allow external facts to count as justifiers. No suitably qualified version of the KJ constraint is derivable from the GD conception. (292)

Nevertheless, Goldman urges that the problems with the accessibility constraint threaten to undermine internalism:

> Internalism is rife with problems. Are they all traceable to the GD rationale? Could internalism be salvaged by switching to a different rationale? A different rationale might help, but most of the problems raised here arise from the knowability constraint. It is unclear exactly which knowability constraint should be associated with internalism, and all of the available candidates generate problematic

theories. So I see no hope for internalism; it does not survive the glare of the spotlight. (293)

What internalism needs, if it is to sidestep Goldman's criticism,[24] is a motivation independent of the accessibility or knowability constraints. At the same time, given the centrality of these constraints in prominent internalist work, one may reasonably wonder whether any such internalism could be fully worthy of the name. A viable internalism, therefore, must not only sidestep Goldman's critique; it must make clear how it is possible to do that while maintaining its integrity as a distinctive philosophical position.

Judged against these standards, how well does my psychologized version of internalism fare? I believe it is clear that this view has the resources to sidestep Goldman's criticism. The view is not committed to any version of the accessibility or knowability constraints. Instead, the view proceeds from a GD conception of justification to a certain view about idealization in epistemology, seeing appropriate epistemic ideals as rooted in human competence. Our ideals, on this view, must have their source *in us* if they are to occupy that large middle ground between ideals so disconnected from our capacities that they lose their ability to guide us and, at the other extreme, ideals so tightly tied to our actual performance as to lose their entitlement to be seen as ideals at all. The ideals are thus internal to us in a straightforward sense. In addition, given the empirical assumptions I have been making on behalf of my psychologized internalist, the substance of the ideals endorsed turns out to be very much like the ideals that paradigm internalists have long championed. Internalism not only can go on without a commitment to the accessibility constraint, it now looks to be proposing an interesting epistemic ideal based on a potentially viable empirical claim. Unlike versions of internalism that seek to defend forms of accessibility or knowability constraints—constraints that are committed to claims about introspection which, I believe, are demonstrably false—a psychologized internalism of the sort I am proposing has some real chance of empirical vindication.

A psychologized internalism thus has far more than a casual kinship with views that everyone would agree are internalist. It presents a worthy rival to current externalist theories of justification, and it shows us one way in which deontological views may find a natural home in epistemology.

IV

I've been suggesting that the psychologized version of internalism stands or falls with a controversial empirical claim: the claim that the mental logicians are right in thinking that we are innately endowed with a mental module hard-wired with a powerful formal system that guides much of our thought. But I want to consider a suggestion on behalf of internalism that seeks to provide it with a less controversial backing, a view that does not require one to take a stand on the issue that divides the mental logicians from those who see our thought as largely guided by domain-specific reasoning strategies.

You and I are perfectly capable of reasoning in accord with formal principles. If you ask me to suppose, for the sake of example, that p, and also to suppose that $p{\rightarrow}q$, I will quickly add q to the list of things that I am supposing and I will make use of it in whatever reasoning I engage in when considering your proposed example. I am thus not only capable of reasoning in accord with formal principles, I actually engage in formal reasoning on many occasions. Now note that this seems very clearly true whatever one may think about the innate principles that underlie my thinking. That I am capable of formal reasoning is an empirical claim, to be sure; but it is an empirical claim that I have excellent reason to believe, and it requires no fancy theorizing or support from cognitive science. Each of us is in a position to recognize that human beings are capable of formal reasoning of the sort that internalists see as ideal, and thus internalists may ground their philosophical ideals in human capacities even without consulting the cognitive sciences. Internalism, it will be argued, should thus be uncontroversial. It is indeed far better confirmed than any current view in the cognitive sciences. This kind of defense of internalism will recognize a grain of truth in the claim of naturalists who have argued that an adequate defense of a normative epistemology must ultimately be empirically based, but, in the end, the empirical basis that this sort of internalist will rely upon is far narrower than naturalists have urged. It is, indeed, an empirical basis that is so obvious that, in practice, epistemological theorizing may go on from the armchair, even if, strictly speaking, the resulting epistemological theory is not defensible in a fully a priori way.[25]

But this defense of an internalist epistemology is unsuccessful. While it is certainly true that human beings are capable of reasoning in accord with formal principles, whatever the ultimate basis of human reasoning, an adequate internalism needs a good deal more than just the principles of, for example, first-order propositional logic. Whether there are formal principles available to do the work that internalists need done is something about which reasonable people might, and do, disagree. A bare-bones evidentialism that makes no attempt to fill in the details of the needed formal principles does not begin to address this issue.

Now there are two different ways in which evidentialists might try to fill in this gap. One is to construct the needed principles themselves: to show exactly how it is that one might justify a large body of claims by appealing to available evidence and the needed formal principles. Certainly this is the sort of thing that, say, Carnap attempted, without success. But another way in which one might defend an evidentialist epistemology, short of coming up with the relevant principles oneself, is less direct: if one had evidence that the human mind actually has the kind of native competence that evidentialists aspire to, then even short of knowing the details of precisely how it is that we are able to make the relevant inferences, one might believe that the evidentialist project is one that can actually be carried out. The work of the mental logicians looks for all the world like an evidentialist project. It is not just that the mental logicians claim that certain kinds of formal principles are hard-wired in the human mind; rather, it is their idea that principles such as these might underlie a substantial part of human reasoning. One might well have evidence for this sort of claim without yet being in a position to give a complete account of the principles by which the mind operates. And if one did have such evidence, one would be in a po-

sition to defend a version of evidentialism. But this, of course, would not rely on the easily available evidence that we all are acquainted with in the armchair; it would require the more abstruse empirical theorizing provided by cognitive science.

In addition, our ability to recognize at least some of the kinds of reasons that internalists favor, and to draw conclusions in accord with them, counts for less than this kind of defense of internalism would have us believe. In particular, suppose that, although we are able, at least locally, to recognize certain formal relations among propositions, this ability is not attributable to operations in the mind in accord with such formal principles, as the mental logicians would suggest. Instead, suppose that this ability is attributable, at bottom, to the operation of domain-specific reasoning strategies, or, instead, the operation of less reliable but more computationally feasible formal strategies. If this were the case, then when we ask what our reasoning mechanisms are truly capable of, we need not be able to generate all that a mechanism guided by the formal principles themselves would generate, and this will remain true even when we idealize so as to ignore interfering factors.[26] Insofar as evidentialism is tied to a theory of epistemic obligation, and thus wishes to ground those obligations in what our reasoning mechanisms themselves are, in some robust sense, genuinely capable of, the question of how our minds actually operate will remain relevant. So long as it does, the quick defense of evidentialism provided from the armchair proves inadequate.

V

In the end, I myself do not favor any version of internalism at all. I have been convinced by the empirical arguments in favor of domain-specific reasoning strategies, and I thus believe that an account of epistemic obligations will provide one more route to externalism. At the same time, I recognize that these empirical arguments are far from conclusive and that reasonable people currently disagree about the force of the available evidence. I have thus tried to spell out a version of internalism that comports with this other reading of the evidence.

I recognize that most internalists will not see this effort on my part as any sort of friendly help. Many of them, Feldman included, wish to hold on to some version of the accessibility requirement. But because I believe that any such view is completely untenable, and because, in spite of that, I believe that there is much in the internalist ideal that might yet be saved, I believe that it is worth attempting to reconstruct an internalist ideal that eschews accessibility. If I am right, this is internalism's best hope.[27]

Notes

1. William Alston, "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives*, 2 (1988): 257–99.
2. Alvin Plantinga, *Warrant: The Current Debate* (New York: Oxford University Press, 1993), chapter 2.

3. "Epistemic Obligations," *Philosophical Perspectives*, 2 (1988): 235–56; "Voluntary Belief and Epistemic Evaluation," in *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility and Virtue*, M. Steup, ed. (New York: Oxford University Press, forthcoming); "The Ethics of Belief," *Philosophy and Phenomenological Research*, forthcoming; and, with Earl Conee, "Evidentialism," *Philosophical Studies*, 48 (1985): 15–34.

4. But see Bruce Brower's "The Epistemic Contract," manuscript, for an attempt to provide just such a contractual account of epistemic obligation.

5. Feldman, "Voluntary Belief and Epistemic Evaluation," manuscript, 19; and also "The Ethics of Belief," manuscript, 7.

6. "Voluntary Belief and Epistemic Evaluation," 20; "The Ethics of Belief," 7.

7. Ibid., 19.

8. Ibid., 20.

9. "The Ethics of Belief," 9.

10. Ibid.

11. Ibid., 11.

12. Elsewhere, Feldman criticizes an account of what it is for an agent to have evidence available for a proposition when he notes, "it seems clear that there could be available evidence whose relevance is unappreciated by a particular believer." The account he there considers "has the weakness that it excludes from the evidence one has relative to a particular proposition other beliefs whose relevance one fails to appreciate." See "Having Evidence," in *Philosophical Analysis: A Defense by Example*, David Austin, ed. (Dordrecht: Kluwer, 1988), 93.

13. Feldman and Conee discuss an example in which Alfred justifiably believes *p*, justifiably believes if *p* then *q*, but then believes *q*, not because he deduces it from these other two beliefs, but rather because he likes the sound of some sentence expressing that proposition. They claim that in this example the individual has the relevant evidence for *q* and is thus justified in believing it by evidentialist standards, even though he did not use the relevant evidence and is thus criticizable for that. See "Evidentialism," 24–25.

14. One may, of course, avoid the regress by denying that apprehension that *E* supports *p* requires belief that *E* supports *p*. It is not at all clear how such a suggestion might be coherently motivated within the evidentialist framework.

15. "Voluntary Belief and Epistemic Evaluation," 19–20.

16. Ibid., 20.

17. This is certainly the view I had in mind in "Justified Belief and Epistemically Responsible Action," *Philosophical Review*, 92 (1983): 33–48, where I made explicit the assumption that we do not have direct voluntary control over our beliefs and yet, at the same time, defended a conception of epistemic responsibility. John Heil has defended a view that is, in relevant respects, quite similar: see his "Doxastic Agency," *Philosophical Studies*, 43 (1983): 355–64 and "Believing What One Ought," *Journal of Philosophy*, 80 (1983): 752–65.

18. See, e.g., Lance Rips, *The Psychology of Proof: Deductive Reasoning in Human Thinking* (Cambridge, Mass.: MIT Press, 1994).

19. Indeed, I've argued elsewhere, following Georges Rey, that the best hope for a defense of the existence of a priori knowledge is to be found in this empirical literature. See Rey's "The Unavailability of What We Mean I: A Reply to Quine, Fodor and LePore," in *Holism: A Consumer Update. Grazer Philosophische Studien*, J. Fodor and E. LePore, eds., 46 (1993): 61–101; and "A Naturalistic a Priori," *Philosophical Studies*, 92 (1998): 25–43; and my "The Impurity of Reason," *Pacific Philosophical Quarterly*, 81 (2000): 67–89. Alvin Goldman has defended a view similar to Rey's: see his "A Priori Warrant and Naturalistic Epistemology," *Philosophical Perspectives*, 13 (1999): 1–28.

20. See, e.g., Jerome Barkow, Leda Cosmides, and John Tooby, eds., *The Adapted Mind:*

Evolutionary Psychology and the Generation of Culture (New York: Oxford University Press, 1992) and Lawrence Hirshfeld and Susan Gelman, eds., *Mapping the Mind: Domain Specificity in Cognition and Culture* (Cambridge: Cambridge University Press, 1994).

21. John Pollock has suggested this. See "Epistemic Norms," *Synthese*, 71 (1987): 61–96; and *Contemporary Theories of Knowledge* (Lanham, Md.: Rowman & Littlefield, 1986); second edition, with Joseph Cruz (Lanham, Md.: Rowman & Littlefield, 1999).

22. *Journal of Philosophy*, 96 (1999): 271–93.

23. I have made a similar argument; see section 5 of my "How Internal Can You Get?," *Synthese*, 74 (1988): 313–27.

24. Some internalists, I expect, will prefer not to sidestep Goldman's criticism, but instead to mount a head-on attack on his critique of the accessibility and knowability constraints. As someone who has long argued against these constraints myself, I believe that this route is extremely unpromising.

25. Feldman himself suggests something like this in "Methodological Naturalism in Epistemology," in *The Blackwell Guide to Epistemology*, J. Greco and *E*. Sosa, eds. (Oxford: Blackwell, 1999), 170–186.

26. For an excellent discussion of these issues, see Christopher Cherniak, *Minimal Rationality* (Cambridge, Mass.: MIT Press, 1986).

27. I have been helped by Guy Axtell, participants at the conference at Santa Barbara, and especially by David Christensen and Richard Feldman.

INDEX